The Impact of Big Data Analytics Maturity on Firm Performance: Evidence from the UK Manufacturing Sector

By:

Deepak Arunachalam

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

The University of Sheffield

Faculty of Social Sciences

School of Management

20 September 2018

Dedicated

to the almighty for providing strength and courage, and

to my family for their support and putting up with me, and

I am influenced by a famous quote

"what you see can be false, what you hear can be false, truth only emerge as a consequence of critical inquiry" – by Thiruvalluvar, to whom I also dedicate this thesis.

ACKNOWLEDGEMENTS

I would like to take this opportunity to thank several people whose contribution is significant to the successful completion of this thesis.

First, I am forever grateful to my supervisors' Dr Andrew Brint, Dr Chantal C. Cantarelli and Dr Niraj Kumar for their exceptional guidance and outstanding support. A special thanks to Dr Niraj Kumar for his continued support and guidance even after relocating to another institution in the UK. Their advice and critical feedback during the research process are precious to me and without their support and guidance this work could have never been possible. Their profound knowledge of the research domain, positive attitude and constructive feedback throughout this journey have great influence on my personal and professional development. No words could describe my thankfulness to my supervisors; for being compassionate and advising me on career opportunities as well.

I am tremendously grateful to the University of Sheffield and the Management School for providing me this excellent opportunity to obtain a PhD degree. I am forever indebted to the financial support offered by the University. The scholarship support has allowed me to focus on the research and complete the journey without any financial burden.

I appreciate the help of research support and administration team at the Management School for their kind assistance throughout the course of my study. In particular, I would like to thank Ms. Mandy Robertson and Mrs Josie Smith for being very friendly and responding to all the queries I had.

I am very thankful to my friends and colleagues in the United Kingdom for their support and making this an enjoyable journey. In particular, I would like to thank Dr Ben Lowe for his valuable advices, Dr Niladri Palit for sharing his knowledge and helping me with my teaching career, and Gabriela for keeping the office environment more lively. Finally, I am obliged to all the academicians and practitioners who have helped me in advancing my thoughts and ideas.

ABSTRACT

Big Data Analytics (BDA) is perceived as one of the most prodigious technologies of the 21st century. However, relatively few studies have demonstrated the positive outcome of developing BDA capabilities. Moreover, although BDA is argued to advance innovation (Tan et al. 2015), the influence of BDA on innovation performance is yet to be confirmed empirically. Studies explaining the underlying mechanism through which BDA influences operational and innovation performances are limited, especially from a BDA maturity and organisational learning perspective. Further, the phenomenon of a digital divide, between SMEs and large organisations, instigated by the adoption of BDA remains unexplored. Therefore, the main purpose of this study is to address these gaps and contribute to the literature by empirically investigating the value of Big Data Analytics capabilities from a maturity perspective, and to explore the digital divide in the context of the UK's manufacturing sector.

A systematic literature review is conducted to identify potential research gaps and to frame the research questions. Drawing on the Resource-Based View, Dynamic Capabilities, and a hierarchy of capabilities perspective, a conceptual model is developed depicting the relationship between the BDA capabilities, the higher-order capabilities such as absorptive capacity, data and information quality and supply chain analytics, and its effect on innovation and operational performance. A survey-based research method is adopted to investigate the model and test the hypotheses. For data collection, an online questionnaire is distributed via Qualtrics to senior executives in UK manufacturing sector. After a rigorous pre-processing of collected data, a sample of 221 responses is estimated to be appropriate for further analysis. This study used two types of data analysis techniques: Structural equation modelling and Cluster analysis. Structural equation modelling is used to examining the causal relationship between the variables in the conceptual model. Whereas, cluster analysis is used to explore the phenomenon of a digital divide between SMEs and large organisations.

The findings of this study indicate that BDA capabilities improves operational and innovation performances. The impact of BDA capabilities on operational performance is partially mediated by absorptive capacity, data and information quality, and supply chain analytics capability. However, the impact of BDA capabilities on innovation performance is only mediated by absorptive capacity. In terms of cluster analysis, the findings indicate the presence of four homogeneous

clusters of organisations with varied levels of BDA capabilities maturity, signifying the reality of digital divide.

This study makes some significant contributions. First, in terms of the contribution to literature, this study synthesised arguments from the Resource-Based View, the Dynamic Capabilities View and the hierarchy of capabilities view to provide a holistic explanation of the relationship between lower-order capabilities, higher-order capabilities and operational and innovation performances. Second, in terms of contribution to practice, this research will help to improve the practitioners' understanding about how BDA capabilities can improve firm performance. The BDA maturity framework developed in this research can be used by the practitioners to assess the current level of BDA capabilities allowing the organisations to determine the areas of improvement. Third, this research provides implications for policy that could be advantageous to SMEs, who are mainly data and information poor.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LISfT OF ABBREVIATIONS

Abbreviations	Full Terms
AA	Advanced Analytics
ACAP	Absorptive Capacity
AIC	Akaike Information Criterion
BD	Big Data
BDA	Big Data Analytics
BDAM	Big Data Analytics Maturity
BIC	Bayesian Information Criterion
BDS	Big Data Skills
CFA	Confirmatory Factor Analysis
DA	Digital Analytics
DDC	Data-Driven Culture
DCV	Dynamic Capabilities View
DG	Data Generation
DIM	Data Integration and Management
DIQ	Data and Information Quality
DV	Data visualisation
EFA	Exploratory Factor Analysis
IoT	Internet of Things
IOS	Inter-Organisational Systems
MLE	Maximum Likelihood Estimation
OP	Operational Performance
RBV	Resource Based View
SCA	Supply Chain Analytics
SCM	Supply Chain Management
SEM	Structural Equation Modelling

Chapter 1 Introduction

1.1 Chapter introduction

In this chapter, the fundamental thoughts that encouraged the initiation of the entire process of this research are discussed. The chapter discusses the motivation for the research interest in Big Data Analytics and highlights the research context of the study. Further, the research issues and the questions addressed in the study are discussed, following which an outline of this thesis is provided.

1.2 Motivation for the research

The concept of Big data is “cultural, technological, and scholarly phenomenon” (Boyd and Crawford 2012, p.663), and it is gaining momentum in the recent years. One of the critical aspects of Big Data Analytics (BDA) is its influence on how data-driven decisions are made and how information is generated and utilised. Organisations that are digital-savvy strategically use Big Data in a variety of ways to create value for the customers (Frizzo-Barker et al. 2016). In the globalised business world, organisations recognise the use of innovative technologies such as BDA for the management of business operations and gain a sustainable competitive advantage. It is widely acknowledged that utilising BDA to manage intra- and inter-organisational activities can provide organisations with economic benefits and competitiveness (Dutta and Bose 2015; McAfee and Brynjolfsson 2012). Moreover, the application of BDA to manage supply chain activities is gaining great interest among industry practitioners. A typical supply chain includes information flow along with material and financial flow (Souza, 2014). It can be argued that a rapid increase in data and information availability can be witnessed these days, instigated by the tremendous increase in the adoption of supply chain technologies such as Enterprise Information Systems (EIS), Barcodes, Global Position Systems (GPS), Radio Frequency Identification Devices (RFID), Sensors, and more recently the Internet of Things (IoT). These supply chain technologies have the potential to generate data at the various stages of business operation. Consequently, organisations are inundated with data and hence motivated to manage and analyse it so as it could improve operational performance (Chae and Olson, 2013). For instance, a global apparel manufacturing company (Li & Fung) has to manage the inflow of more than 100 gigabytes of data every day (The Economist, 2010). The emergence of such Big Data (BD) from day-to-day operations creates both opportunities and challenges. Optimistically, it is argued that the Third Industrial Revolution (TIR) would be governed by four main components which include Big Data Analytics (BDA), adaptive services, digital manufacturing and mass customization (Tien, 2015). BDA is the use of advanced data management and analytics tools to manage information flow and provide data-driven insights for decision makers. Business leaders prefer data-driven insights for decision-making rather than those based on intuitions (Davenport, 2006). BDA certainly could improve organisations' capabilities to respond to a rapidly changing market environment (Meredith et al. 2012), and firms consider BDA as a strategic asset to effectively manage supply chains. However, due to the complexity of BD technologies, it is obligatory for industry practitioners to tackle some technical and organisational challenges during implementation. It is evident from the literature (Fosso Wamba et al. 2015; Kambatla et al. 2014; Tsai et al. 2015) that BD related technological developments such as Hadoop, NoSQL and MapReduce are growing at a fast rate, and will continue to evolve. Nowadays, the awareness of the benefits of BDA has increased and the rate of adoption could have improved as well. However, limited prominence is given to the behavioural issues associated with the adoption and practice of BDA, and researchers have argued that most organisations are still in the learning phase (Halter and Rao, 2013).

Consequently, this study begins with an idea to examine the importance of Big Data Analytics (BDA) practice for improving firm performance. After a brief review of the literature, it is noticed that the current literature did not contain enough empirical works related to this domain. Moreover, it is recognised that there is a growing interest among academics and industry practitioners to understand how important Big Data Analytics is to improve the performance of an organisation. Subsequently, a systematic literature review is conducted to examine in detail the existing contributions in the literature related to this domain. Adoption of Big Data Analytics and how it can improve organisational learning and performance is thus found to be an important research idea to investigate. Further, perceiving the adoption of BDA technology from maturity (state of being complete) and organisational learning (ability to absorb information and knowledge) perspectives could give a comprehensive view of the phenomenon. So, this research is intended to examine the impact of BDA capabilities from a maturity perspective on Absorptive Capacity (ACAP), Data and Information Quality (DIQ), Supply Chain Analytics (SCA) capabilities and firm performance. Consequently, Systematic Literature Reviews

(SLR) are used to ascertain the dimensions of BDA Maturity (BDAM), ACAP, and operational and innovation performance. Key research gaps are also identified (discussed in section 2.9.1) and it is found that very few empirical studies have investigated the impact of BDA capabilities on operational and innovation performance. Moreover, an important question remains unexplored, which is the underlying mechanism through which the adoption and use of BDA capabilities contribute to firm performance (Chen et al. 2015a; Cao et al. 2015). Despite its popularity within the industry, the reluctance to invest in BDA is also noticeable due to the ambiguity in recognising the potential benefits of BDA. Above all, research on BDA in the context of the UK manufacturing sector is scarce and this nascent domain deserves comprehensive investigation. Therefore, in the following section, the research problems addressed in this study are discussed. After this, the research context, research questions and objectives are discussed in the following sections.

1.3 Research problem

From a technical perspective, the main issue with BD is its complexity. BD creates three main technological challenges such as: 1) collecting and integrating data from disparate data sources distributed in heterogeneous systems, 2) storing and managing heterogeneous data types alongside maintaining efficiency in terms of retrieval, privacy issues and scalability, 3) analysing data streams in real time to perform various operations on data such as prediction, modelling and visualisation to support decision-making (Hu et al. 2014). The dynamic nature of manufacturing businesses means that organisations must face most of the above-mentioned challenges. For instance, in a typical manufacturing supply chain, data is generated in heterogeneous information systems located across Intra- and Inter-organisational boundaries, and are possibly of varied types (structured, unstructured and semi-structured) subject to the formats and maturity level of information systems used. For SCM, data has to be integrated and analysed in real time to effectively monitor processes across the supply chain network, requiring a high level of organisational collaboration and technologies like in-memory analytics (Hahn and Packowski, 2015). The traditional Relational Database Management Systems (RDBMS) can only store structured data and has issues with scalability and storage of semi-structured and unstructured data and are not sufficient to tackle these BD challenges. Past literature such as Wamba et al. (2017) and Chae et al. (2013) claimed that the successful implementation of BDA would optimise the business processes and have a positive impact on firm performance. However, Sherer (2005 as cited in Oliveira et al. 2012) argued that there are several pieces of evidence of BDA failures. In particular, Oliveira et al. (2012) have elaborated on the case of BDA failure in Cisco, a High-tech company, which had faced severe expenditure due to erroneous demand predictions. Similarly, Google's flu trends is another well-known case of failure of BDA (Lazer et al. 2014). In addition, Knight Capital Group lost USD 440 million in 2012 within a few hours after implementing a decision based on malfunctioning automated analytic algorithms (OECD 2015). So along with the benefits, the underlying risks involved in data and analytics algorithms leading to an unexpected false result, are also high. Unlike the claims by the BDA technology vendors, inappropriate decisions from BDA systems can create social and economic harm. In these uncertain scenarios, very little is known about the desired BDA capabilities and the mechanism through which these disruptive technologies can be used effectively in line with organisations' business strategy.

In line with the above discussion, this research attempts to address the following issues related to BDA:

1) In the information era, an organisation's success relies on how proficient they are at taking advantage of their information resources (Olszak, 2016). However, the practice of BDA is not well articulated within the domain of supply chain (SC) research. Understanding BDA capabilities is crucial considering the emerging belief that BDA assets can provide a sustainable competitive advantage. Therefore, since relatively little is known about the different elements of BDA capabilities which are significant to improving performance, there is a *necessity to explore the multi-dimensions of BDA capabilities* to create value for the organisation and understand its influence on the performance.

2) Despite the awareness about BDA and its benefits, there exists a causal uncertainty, which acts as a barrier for utilising valuable resources desirable to attain competitive advantage (Reed & Defillippi, 1990). Organisations having a non-imitable resource can create competitive advantage. Reed and Defillippi (1990) argued that the complexity of technological systems is one of the sources of ambiguity, making it difficult for firms to recognise the roots of success and failure. Therefore, it

is necessary to address the *issue of causal uncertainty by explaining the underlying mechanism* through which the BDA capabilities improve firm performance.

3) Another important issue addressed in this study is, *the presence of the digital divide due to the disparity of BDA adoption between large organisations and SMEs*. According to Organization for Economic Co-operation and Development (OECD, p.5) the term digital divide refers to “the gap between individuals, households, businesses and geographic areas at different socio-economic levels with regard both to their opportunities to access Information And Communication Technologies (ICT) and to their use of the Internet for a wide variety of activities’’. The digital divide is the term used to represent socio-economic differences in the usage pattern of ICT (Vehovar et al. 2006). However, in the context of this research, the focus is on the adoption of BDA in organisations. In a study, it is revealed that SMEs in the UK lack knowledge about BDA and its benefits (SAS, 2013). The proportion of SMEs in the UK that have utilised BDA is found to be negligible, with only 0.2 per cent of SMEs participating in the study showing awareness of it (SAS, 2013). The disparity in the adoption trend of these innovative technologies would certainly hinder the growth of SMEs in the UK, and as a result, UK SMEs would lack a competitive advantage in the global marketplace. Tien (2012) has mentioned the Data Rich and Information Poor (DRIP) conundrum - referring to the state of having an enormous volume of data but less information available for decision-making. In the organisational context, it can be argued that large organisations do not have restrictions in terms of economic capital and technical knowledge for the adoption of new technology. However, SMEs could certainly fall into the category of Data Poor and Information Poor (DPIP) (Figure 2.12) as it can be relatively difficult for them to harness BDA. Therefore, it is necessary to investigate the level of BDA maturity for the following reasons: 1) to gather evidence by comparing the differences in BDA maturity level between SMEs and large organisations, 2) to help SMEs evaluate their current state using the maturity model and recognise the capabilities they need to implement to compete with large organisations, and 3) to support policy-makers and technology vendors to develop policies, customised products, and services to promote SME growth. So, in order to address these problems, this research aims; 1. to examine the casual relationship between Big Data Analytics maturity and firm performance dimension, and 2. to explore the phenomenon of digital divide caused by the adoption of BDA technology.

1.4 Research context

The main focus of this research is the BDA practice in the context of the UK manufacturing sector. A brief overview of the concept of the supply chain, supply chain management, and the rationale for selecting the UK manufacturing sector is discussed here.

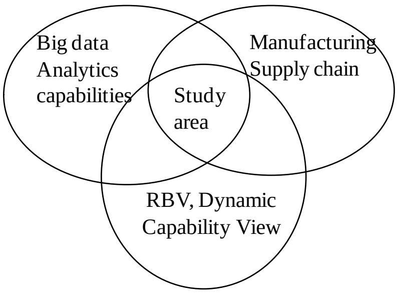


Figure.1.1 Context of the research

According to Mentzer et al. (2001, p.4), the supply chain is

"a set of three or more entities (organisations or individuals) directly involved in the upstream and downstream flows of products, services, finances, and/or information from a source to a customer".

Supply chain managers are responsible for this process. The definition encompasses the aspect of information flow between organisations in the supply chain, which is the prime focus of this research. Further, SCM is defined as

"the management of upstream and downstream relationships with suppliers and customers in order to deliver superior customer value at low cost to the supply chain as a whole" (Christopher, 2016, p.2).

The intention to implement BDA in SCM to manage the information flow is increasing (Wang et al. 2016a), and companies are inclined to use BDA to improve forecasting accuracy, warehouse optimisation, reducing costs, visibility, and integrating supply chain processes (Genpact, 2015). As discussed by Wang et al. (2016) and Sanders (2014), BDA is composed of two separate components: 1) Big Data, more complex than traditional data sets, and 2) Analytics, in a broader sense, comprises techniques such as maths, statistics, optimisations, and simulations to extract insights from BD. Davenport and Harris (2007, p.7) defined the term analytics as *"the extensive use of data, statistical and quantitative analysis, explanatory and predictive models, and fact-based management to drive decisions and actions"*. So, neither of these components are useful when taken separately (Sanders, 2014). The combination of BD & Analytics makes the difference and has emerged as an imperative technology to turn data into actionable insights. The discipline of BDA in SCM has drawn attention because of its pivotal role in improving firm performance and the necessity to address several issues such as casual ambiguity.

Manufacturing organisations have changed their strategy in response to the changing market demands and have become more customer-centric (Aho, 2015). They increasingly consider embedding services into products as a differentiation strategy. The transformation of manufacturing firms from the traditional product offering to servitization involves developing new capabilities such as BDA capabilities. Moreover, three major obstacles faced by the executives in manufacturing organisations while they strive to improve operational activities: inundation, isolation, and indecision of information (Jain et al. 2011). Dutta and Bose (2015) have argued that compared to B2B, B2C sectors such as retail have a lot of potential for generating a huge volume of data, and since the manufacturing sector operates mainly in the B2B environment the data generated will be considerably less than in B2C sectors. However, in the near future, it is anticipated that due to the adoption of sensors and the IoTs in the manufacturing sector, there is potential for an explosion of manufacturing data.

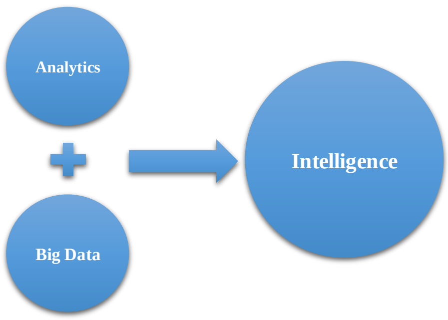


Figure 1.2 Turning information into intelligence

Source: (Sanders, 2014)

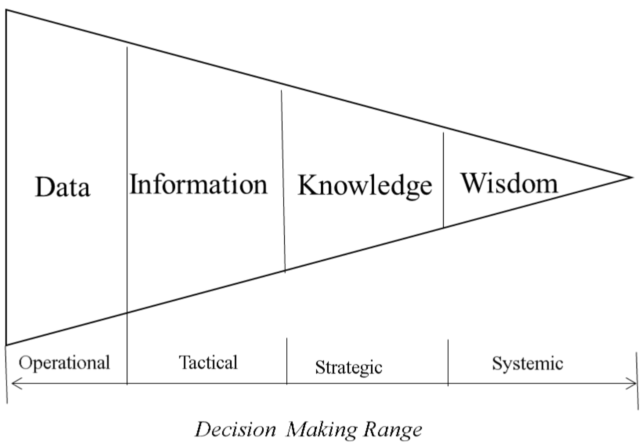


Figure 1.3 Decision-making framework

Source: (Tien 2012, 2015)

Moreover, in the current turbulent economic situation, the manufacturing sector is facing a drastic change due to industry 4.0, advanced technological evolution and adoption behaviour (Almada-Lobo, 2016; Caputo et al. 2016). According to BIS (2015), the UK has 275,565 manufacturing businesses with a combined annual turnover of £ 601 billion. As per the statistics, the majority of UK manufacturing businesses are SMEs with fewer than 250 employees. Large manufacturing organisations represent only 0.4 % of the industry, but they account for 42.5% of total employment. However, small (0-49 employees) and medium (50-249 employees) size organisations' contribution to annual turnover is 12.3% and 16.4% respectively, which is comparatively low compared to larger organisations who contribute 69.8 % of the annual turnover. UK's manufacturing productivity has slumped since the recession in 2009 and has not reverted back to the pre-crisis situation (Hardie and Banks, 2014). Nevertheless, the UK has been specialising in high-tech manufacturing industries to compete against developing economies such as BRIC nations (Brazil, Russia, India, and China), who still focus on low technology manufacturing (BIS, 2015). The UK's strength and focus on high-value manufacturing industries such as aerospace, power generation, and automotive industries has the potential for developing innovative products and services (Hague et al. 2016), and is expected to yield better competitive advantage in international markets (BIS, 2010b). Under this current scenario, companies need better ICT to effectively manage global supply chain networks. As shown in Figure 1.4, the diffusion of manufacturing technologies is progressing at a fast rate, and manufacturing firms rely on the latest ICT technologies such as sensors, RFID, and IoT to manage their business operations. As a consequence, the manufacturing companies are flooded with BD to a larger extent than any other business sector (Zhong et al. 2015, 2016), which necessitates implementing BDA to manage the data flow and extract insights from it. Therefore, understanding BDA capabilities in this context is more significant for the development of SMEs and the UK economy.

Table 1.1 Number of manufacturing companies in the UK

Source: (BIS, 2015)

Number of Employees	Number of manufacturing companies in the UK
With no employees (unregistered)	146,030
With no employees (registered)	41,565
1	6,370
2-4	33,365
5-9	18,865
10-19	12,955
20-49	9,230
50-99	3,720
100-199	1,825
200-249	400
250-499	710
500 or more	530

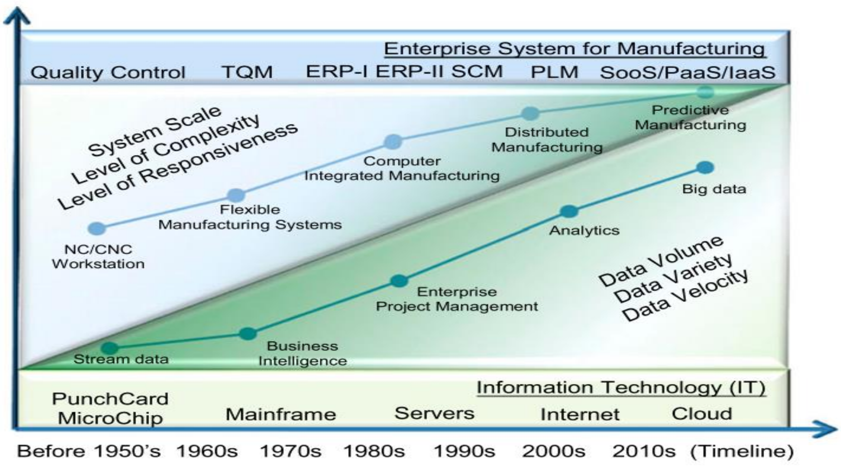


Figure 1.4 Evolution of Manufacturing Enterprise systems

Source: (Bi and Cochran 2014)

1.5 Research questions

The two fundamental research questions addressed in this study are:

RQ1. What is the impact of Big Data Analytics Maturity on firm performance in the context of the UK manufacturing sector?

In relation to the main research question (RQ1) the following sub-questions are addressed:

> RQ1a. What is the relationship between Big Data Analytics capability maturity and firm performance dimensions?
>
> RQ1b. What is the role of Absorptive Capacity on the relationship between BDA capability maturity and firm performance?
>
> RQ1c. What is the role of Data and Information Quality (DIQ) on the relationship between BDA capability maturity and firm performance?
>
> RQ1d. What is the role of Supply Chain Analytics capability (SCA) on the relationship between BDA capability maturity and firm performance?

RQ2. To what extent BDA adoption extend the digital divide between SMEs and Large organisations in the UK?

1.6 Research significance

This study aims to make significant contributions to literature, practice, and policy. As discussed in the previous section, the aim of this research is twofold: first, to gain more insights into the impact BDA capabilities have on the performance of the UK manufacturing firms, and second, to investigate the phenomenon of the digital divide. By holistically applying RBV, DC view and the view of the hierarchy of capabilities, this research proposed a conceptual model in which BDA capabilities maturity, as lower-order capability, exert influence on operational and innovation performance through higher-order capabilities, such as ACAP, DIQ, and SCA capability. Through achieving these research aims; this study's academic significance is as follows:

- This research is one of the few academic works that provides a comprehensive view of BDA capabilities (intra- and inter-organisational BDA capabilities) informed by theory and practice.
- This research contributes to validating the assumptions of Resource-based View (RBV), Dynamic capabilities view (DCV), and hierarchy of capabilities view within the context of BDA practice.

- This study empirically clarifies the importance of BDA capabilities within manufacturing context in improving operational and innovation performance. The research contributes to the literature by improving the understanding of BDA practice in manufacturing industries. This study participated in developing the dimensions of BDA capabilities maturity which would contribute to future research studies.
- The study provides a conceptual model for BDA capabilities as a source of sustainable competitive advantage. This model is higher-order in nature consisting of multidimensional constructs and might be useful in informing further research related to BDA practice.
- The study enhances the understanding over the current state of SMEs and large organisations relating to the adoption and practice of BDA capabilities, which would contribute to the literature on SMEs growth.

1.7 Thesis structure

The entire thesis is structured into seven distinct chapters as shown in Figure 1.5. A brief overview of the remaining chapters is presented below.

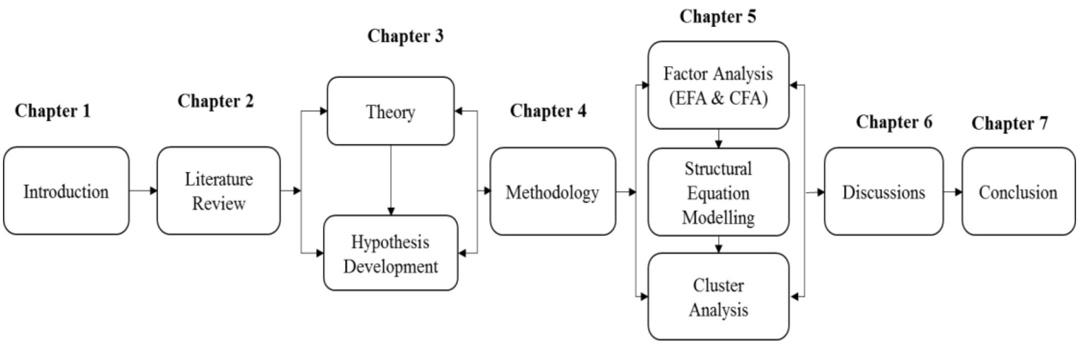


Figure 1.5 The structure of this thesis

Chapter 2 This chapter reviews the existing literature on Big data analytics, absorptive capacity and operational and innovation performance to find potential research gaps and to conceptualise the dimensions of BDA. To conduct a literature review, a scientifically rigorous systematic literature review methodology is adopted. The review presents the research agenda and the scope of examining the relationship between BDA maturity and firm performance dimensions and further identifies the importance of absorptive capacity, data and information quality, and supply chain analytics capabilities.

Chapter 3 Drawing on the existing studies and underpinned by the Resource-Based and Dynamic-Capabilities view, this chapter develops a conceptual model and articulates several hypotheses to explain the role of lower-order and higher-order capabilities in improving the operational and innovation performance.

Chapter 4 This chapter covers the transformation process of the research. A transformation from purely theoretical to a philosophical and methodological perspective is presented in this chapter. Consequently, this chapter illustrates the key decisions made underpinned by the philosophical and methodological considerations. Based on the chosen approach, the mono-method quantitative research strategy is explained. This chapter also discusses the rationale for the data collection and analysis techniques used in this research. Further, the operationalisation process of the constructs measured in this research is also elaborated.

Chapter 5 This chapter presents the quantitative investigation of the data collected using an online questionnaire survey. Two main data analysis techniques (Structural Equation Modelling and Cluster Analysis) used in this research are presented in this chapter. Prior to the main data analysis, the chapter examines the suitability of the data and its structural characteristics. This chapter also presents the exploratory and confirmatory data analysis used to ensure the validity of the scales used to measure the constructs. After the factor analyses, the refinement of the initial conceptual model is articulated. Then, the implementation and results of Structural Equation Modelling used to examine the hypothesised relationship between BDA maturity, absorptive capacity, data and information quality, supply chain analytics capabilities and operational and innovation performance is presented. Finally, this chapter also includes details of cluster analysis performed to explore the current state of SMEs and large organisation and verify the digital divide between them. Finally, this chapter summarises the key findings of the statistical tests used in this research.

Chapter 6 This chapter discusses the findings from the analysis of the survey data. The analysis techniques used in the previous chapter are aimed at testing the hypothesised direct and mediation relationships of the conceptual model, as well as recognising the digital divide. This chapter draws on the findings of each hypothesis test and discusses its relevance to the context of this research. Further, the optimal number of clusters identified by comparing various clustering algorithms are characterised and presented in this chapter. This chapter also draws on the findings to provide some managerial implications for the best practice of BDA

Chapter 7 Finally, this chapter concludes the findings of this research. It discusses contributions to the theory, practice, and policy. Limitations of this research and suggestions for future research are also provided.

1.8 Chapter Summary

In summary, this chapter presents a brief outline of the motivation, the research problem and the relevance of this research in understating the BDA practice and its influence on firm performance. The objective of this chapter is to provide the rationales for the research questions, the research context and the significance of the study. This chapter also outlined the structure of the thesis to clearly navigate through different chapters of the thesis. The next chapter will focus on the review of the literature to concentrate on the study's scope and to identify potential research gaps.

Chapter 2 Literature review

2.1 Chapter introduction

This chapter defines the concept of Big Data analytics, its key capabilities and how the technology has evolved. Moreover, it reviews mechanisms governing how manufacturing firms can extract full value from their BDA investment (i.e., absorptive capacity, data and information quality, supply chain analytics capabilities). This chapter covers the Systematic Literature Review (SLR) methodology utilised to assess the previous literature related to the domain of: 1. BDA and supply chain management, 2. Absorptive capacity and supply chain management. Consequently, this chapter presents the findings of the two Systematic Literature Reviews (SLR), the methodology used to conduct SLR and the important gaps identified because of it. This chapter also presents the literature review of performance outcomes such as operational and innovation performance.

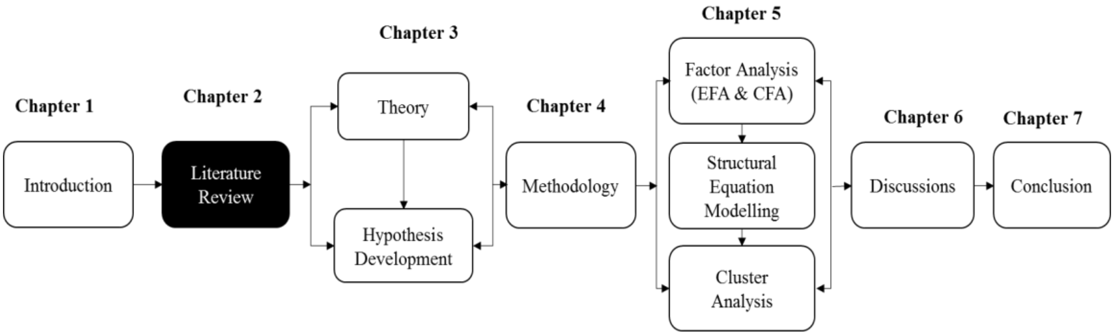


Figure 2.1 The position of the chapter in this thesis

2.2 Business Intelligence (BI), Business Analytics (BA) and Big Data Analytics (BDA)

Data, according to the Oxford English Dictionary (OED), is defined as "facts and statistics collected together for reference or analysis" or "the quantities, characters, or symbols on which operations are performed by a computer, which may be stored and transmitted in the form of electrical signals and recorded on magnetic, optical, or mechanical recording media." (Oxford Dictionary, 2016). In line with the first definition, data are mainly facts and numbers collected for further analysis. The analysis of such data can be performed with little effort using traditional statistical methods and mathematical tools without the need for a computer. Whereas, the second definition encompasses 'characters' and 'symbols', emphasising the usage of the computer to perform analysis, and also the utilisation of additional medium to transfer and store data in various forms. Similarly, the term 'Big data' is defined by OED as "an extremely large data sets that may be analysed computationally to reveal patterns, trends, and associations, especially relating to human behaviour and interactions." These definitions have simplified the delineation and comparison of traditional data analysis with big data analysis. BDA uses more sophisticated computational techniques to handle complex data that has been increasing on a large scale, and are unable to be processed using traditional methods.

However, until recently, Business Intelligence (BI) is considered as an umbrella term which includes tools, techniques, and activities that converts raw data into useful information and support decision-making at different levels (operational, tactical, and strategic) (Gudfinnsson et al. 2015). According to Sahay and Ranjan (2008), "BI refers to the use of technology to collect and effectively use the information to improve business potency." Decision support, statistical analysis, data mining, forecasting, and OLAP are the key capabilities of BI, and four main components of BI are data sources, data marts, data warehouse and query and reporting tools (Sahay and Ranjan 2008). Data warehouse is central to any BI solution; data from internal and external sources are extracted and loaded into the data warehouse (Gudfinnsson et al. 2015). Another terminology that has been in practice is Business Analytics (BA). Business users consider 'Business Analytics' as essential for providing data, information, and knowledge to support decision making (Chen et al. 2012; Acito and Khatri 2014; Laursen and Thorlund 2010), and its scope supposedly extends beyond traditional BI reporting. In the literature, BI and Business Analytics (BA) terms are often used interchangeably. Chae et al. (2014), referred BA as "the application of a broad range of analytical techniques and methods and data-driven analytic methodologies to different business domains." However, Laursen and Thorlund (2010, p.12) defined BA as "delivering the right decision support to the right people at the right time." Further, a related term called 'Decision Support Systems (DSS),' which arguably emerged during 1970's is also widely used in this context to indicate the use of technological solutions to support decision-making problems (Shim et al. 2002). Bartlett (2013, as cited in Mortenson et al. (2015)) considers BI as an amalgamation of Business Analytics and Information Technology. Gudfinnsson et al. (2015) also supported the argument of considering Business analytics (BA) as an integral part of BI. However, several researchers have a contrasting opinion and argued that BI is a division of analytics, and a new acronym 'BI&A' as a composite term is used (Chen et al. 2012). Mortenson et al. (2015) argued that there is a similarity between 'analytics' and 'Operations Research' or 'Management Science' as these concepts are related to improving operations and business decision-making.

Furthermore, in the IT literature, scholars have perceived the use of ICT at different levels such as intra-organisational systems and inter-organisational systems (Zhang et al. 2016). Savitskie (2007) argued that the practice of information technology for information sharing within a firm can be denoted an intra-organisational system. On the other hand, Bhakoo and Choi 2013 (p.432) defined Inter-Organisations Systems (IOS) as "the technology-based infrastructure that acts as a conduit for facilitating transactions, sharing information with trading partners, coordinating activities and establishing governance structures between firms. Similarly, BDA practice can be categorised into intra- and inter-organisational BDA systems. In this regard, Supply Chain Analytics (SCA) is a context-specific term, used to denote inter-organisational practices, indicating Big Data and Analytics activities in supply chain management (Wang et al. 2016; Sahay and Ranjan 2008; Souza 2014). Chae et al. (2014) argued that there are three sets of resources (Data management resources, IT-based supply chain planning resources, and Performance management resources) collectively constitute SCA. In general, supply chain analytics is the use of information and analytics tools to support efficient flow of material along the supply chain. Various definitions of SCA can be found in Rozados and Tjahjono (2014).

Although there are different terminologies used in the literature, they all focus on one objective, i.e., extracting value from data. In addition, there is a pattern of evolution regarding the terminologies and development of capabilities suitable for data-driven decision-making. Chen et al. (2012) argued that the concept of first-generation BI (in the 1990s) had evolved as a consequence of advanced statistical techniques in 1970s and data mining techniques in the 1980s. While data mining has been available for decades, it has only recently been commercially accepted due to the data inundation problem and technological developments (Stefanovic, 2015). The second generation BI 2.0 evolved during the early 2000s (the period also witnessed the growth of the internet and web-based systems). Subsequently, BI 3.0 emerged in the 2010s, leading to the era of Big Data. The term 'Big Data' was primarily used by Cox and Ellsworth (1997) to refer to the storage challenges of datasets that are quite large and demand additional resources. The main distinction between traditional BI solutions and Big Data (BD) technologies is database scalability and the ability to store a variety of data types (structured and unstructured) in real-time. Based on the literature, the evolution of BDA is illustrated in Figure 2.2. Hence, while BDA is not new, it has certainly evolved to meet the changing information processing needs of organisations. From 1950 to 2010, the complexity of data has increased gradually, and as a result, BDA has emerged as a flagship technology to tackle BD challenges.

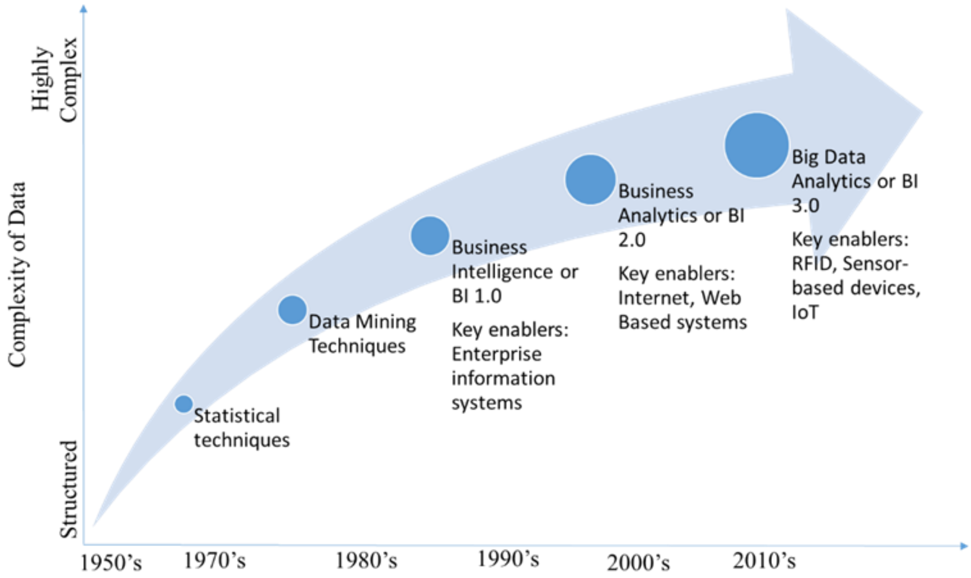


Figure 2.2 Evolution of Big Data Analytics

A range of definitions of 'Big data' can be found in Wamba et al. (2015). A widely accepted definition distinguishing between the volume, variety and velocity of BD from Gartner is given below:

"*high-volume, -velocity and -variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making*" (Beyer and Laney, 2012, p.2).

BD is often related to technological advancement and associated with new kinds of database architectures such as Hadoop, NoSQL, and distributed parallel processing of data. BD technologies are not only able to handle the volume of data but also can effectively manage a variety of data types such as textual information from online blogs, customer reviews, etc. (Mortenson et al. 2015). Vera-Baquero et al. (2015) argued that traditional BI systems are not process aware and are insufficient to integrate data from heterogeneous data sources. In addition, according to Manyika et al. (2011), "Big data refers to datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyse.” Based on the nature of data, the BD is characterised mainly by three dimensions such as ‘Volume,’ ‘Velocity,’ and ‘Variety’ (Manyika et al. 2011; Sonka, 2014). However, apart from the 3V’s (Figure 2.3), BD can also be characterised by another two dimensions ‘Veracity’ and ‘Value’ (Manyika et al. 2011; Neaga et al. 2015; Ge and Jackson, 2014). ‘Volume’ refers to the magnitude of data generated; ‘Variety’ refers to “structural heterogeneity in a dataset” (Gandomi and Haider 2015, p.138); ‘Velocity’ refers to the speed at which data is generated, analysed and acted upon (Gandomi and Haider, 2015); ‘Veracity’ or Verification refers to ensuring data quality, verifying unreliable and uncertain data; and ‘Value’ relates to the economic benefits of Big Data (Mishra et al. 2016). Further, Hofmann (2015) defined the three V’s of Big Data from capabilities perspective as “the firm’s ability to successfully process a large volume of data, integrate various sources of data and process at a high speed (p.3)”. Further, from the capabilities perspective, BDA is defined as “the capability to manage and analyse petabytes of data enable companies to deal with clusters of information that could have an impact on the business” (Hurwitz et al. 2013, p.22). Wang et al. (2016 b) defined BDA from an information lifecycle management view in the context of health care as “*the ability to acquire, store, process and analyse a large amount of health data in various forms, and deliver meaningful information to users that allow them to discover business values and insights in a timely fashion* (p.4)”.

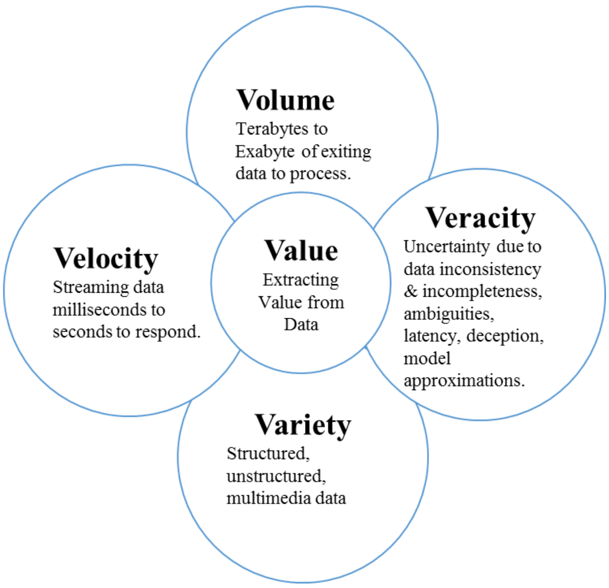


Figure 2.3 Five V's of Big Data

Thus, this section provided an overview BDA and how it has evolved. However, BDA practice is a complex phenomenon with potential to significantly improve the information processing capability of an organisation. In order to comprehensively understand various of aspects of BDA capabilities and its relation to the concept of Absorptive Capacity (ACAP), two systematic literature reviews were conducted as part of the research. First, a systematic literature review of BDA in the context of Supply Chain Management (SCM) was conducted to conceptualise the key dimensions of BDA capabilities. The methodology used and the findings of the first SLR is presented in section 2.3. Subsequently, another systematic literature review of ACAP in the context of SCM was conducted to summarise existing studies and recognise how it has been used in the context of organisational learning, and the findings of the second SLR is presented in the section 2.6.

2.3 Systematic literature review I - Big Data Analytics

The concept of a Systematic Literature Review (SLR) originated in the medical science discipline (Durach et al. 2017), but it is widely recognised in SCM because of its evidence-based approach (Banomyong et al. 2017; Colicchia and Strozzi, 2012). SLR is a positivistic approach and an effective tool to assess an existing body of knowledge and develop it further (Tranfield et al. 2003). It is claimed that SLRs focus on very specific research questions and have the tendency to reduce potential bias (Banomyong et al. 2017).

This section discusses the methodology used to conduct an SLR on Big Data and supply chain management. Also, it presents the descriptive findings and the conceptualisation of BDA capabilities based on the thematic analysis. While there are a few literature review papers (Mishra et al. (2016), Wamba et al. (2015), Donovan et al. (2015), Wamba and Akter (2015)) linking BDA and SCM, these reviews have not discussed BDA capabilities in a manufacturing supply chain context. Also, academic research and reviews related to BDA maturity models are scarce, and therefore this research seeks to address these missing links. This review was aimedto summarise and describe existing research and conceptualise dimensions of BDA capabilities by synthesising the content of current literature. This research follows the literature review approach proposed by Mayring (2003). A similar approach is used by Gao et al. (2016) and Seuring and Müller (2008) in analysing past research papers. This review approach includes four sequential steps: material collection, descriptive analysis, category selection and material evaluation.

2.3.1 Literature search

Section 2.2 demystified the concept of BDA and recognised several terms such as Big Data, Business analytics (BA), Business Intelligence (BI) and supply chain analytics. These terms are used as keywords to search the literature on Big Data Analytics in SCM. Unlike previous Systematic Literature Reviews (SLRs) in this domain, this study adopts a holistic approach by including all possible terms related to BDA practice in businesses. Similarly, different keywords related to SCM are also identified. Different combinations of terms were used to search relevant research papers. Scopus and Web of Science (WoS) databases were used to search related peer-reviewed papers for review. Table 2.1 summarises the keywords used for the literature search along with the number of papers retrieved during the initial search.

2.3.1.1 Inclusion and exclusion criteria

First, only journal papers published in the English language were included. Consistent with Fahimnia et al. (2015), conference papers, papers in commercial magazines, and book chapters were excluded from the search to ensure quality, and only journal papers, reviews, and papers in the press were included. Although the initial search was not restricted by a time limit, the final shortlisted papers are published in the years between 2008 to 2016. The variety of search strings used in the SLR process is given in Table 2.1. The initial shortlisting produced 619 papers. After removing the duplications and verifying the list in Endnote software, the full text of the remaining papers was read to further eliminate irrelevant papers. Only papers which clearly describe the application of BDA was selected for review. This finally resulted in a total of 82 papers spanning from 2008 to 2016. Next, the content of the selected papers was reviewed and classified based on categories such as the distribution of publication year, research methodology, among others. The analysis/evaluation process was complemented by the use of bibliometric analysis - to summarise existing research, and thematic analysis - to conceptualise the content of literature. For the bibliometric analysis, BibExcel Software was used that requires meta-data information of selected journal papers in RIS format, which was extracted using the Scopus database. The findings from the descriptive analysis are given in section 2.3.2. Further, the contents of each shortlisted journal paper were analysed to find themes, codes and conceptualise dimensions of BDA capabilities maturity. The findings of the thematic analysis are presented in section 2.3.3. In addition to the 82 papers from academic journals, for conceptualisation purpose, 13 maturity models sourced from both academic and non-academic sources such as Gartner and Informs were included in the review process. The list of 82 papers and the review of 13 maturity models are given in Appendices F & G.

Table 2.1 Initial search results

Search terms	SCOPUS	WoS
"Big Data" and "Supply chain"	104	65
"Big Data" and "logistics"	101	30
"Big Data" and "operations management"	15	8
"Big Data" and "manufacturing"	20	13
"Big Data" and "operations research"	6	3
"Business Analytics" and "supply chain"	10	13
"Business Analytics " and "logistics"	6	4
"Business Analytics" and "operational performance"	2	2
"Business Analytics" and "operations management"	3	3
"Business Analytics" and "operations research"	4	5
"Business Intelligence" and "supply chain"	64	35
"Business Intelligence" and "logistics"	39	17
"Business Intelligence" and "operational performance"	7	4
"Business Intelligence" and "operations management"	3	3
"Business Intelligence" and "operations research"	6	3
"Business Intelligence" and "operational performance"	7	4
"Supply chain analytics"	8	4
"Supply chain" and "predictive analytics"	16	13

2.3.2 Descriptive analysis

2.3.2.1 Distribution of papers per year

Findings suggests that there is an increasing trend in terms of a number of papers published in the field of BDA in the supply chain (Figure 2.4). In particular, a number of papers were published during 2015 and 2016 compared to previous years. Among the 82 selected papers, most were published in the last four years. The trend shows that there is growing importance among researchers to investigate the phenomenon of BDA in the supply chain context.

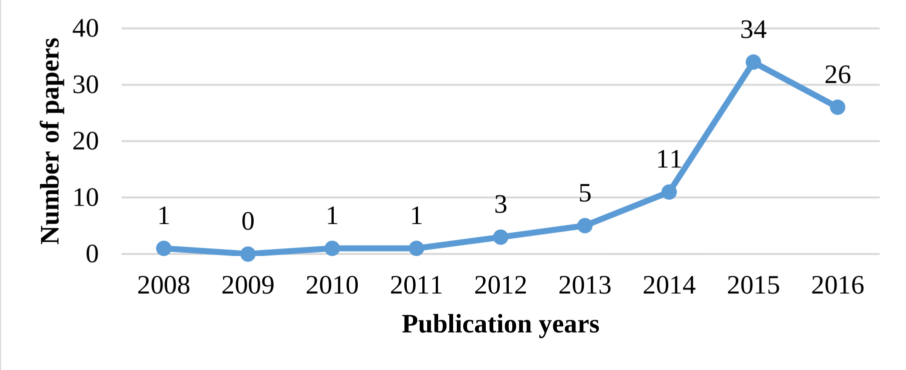


Figure 2.4 Distribution of research papers over the years

2.3.2.2 Top contributing authors

The top 10 contributing authors were extracted using BibExcel tool (Figure 2.5). It is evident that Gunasekaran tops the list with 7 publications followed by Childe, Huang, Hazen, Papadopoulos, Wamba, and Zhong. Moreover, the contributions of authors were further evaluated using h-index and citation counts. It was found that Fawcett and Waller dominate on these criteria, followed by Chae and Gunasekaran (see Figure 2.5 & Figure 2.6).

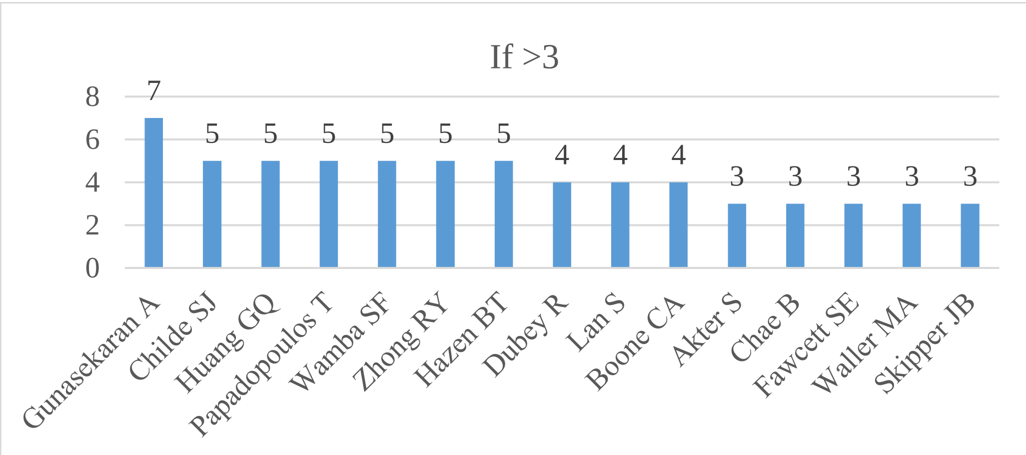


Figure 2.5 Top contributing authors based on the number of publications

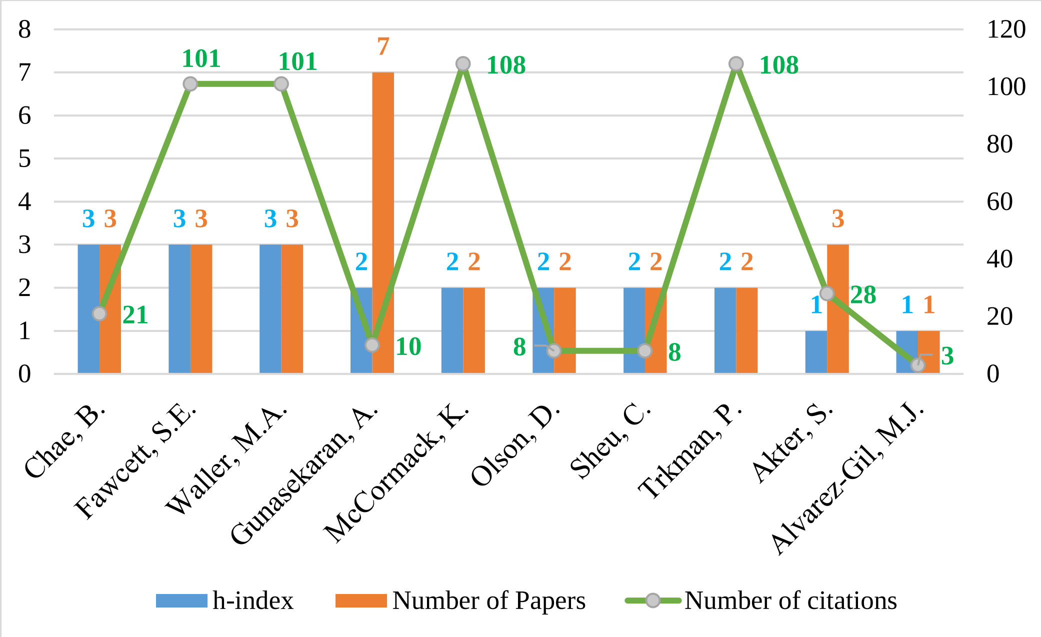


Figure 2.6 Top contributing authors based on h-index, number of publications and citations

2.3.2.3 Affiliation statistics

The affiliations of all first authors were extracted using BibExcel. The United States dominated the top 5 list of contributing countries, followed by China, the United Kingdom, India, and Germany.

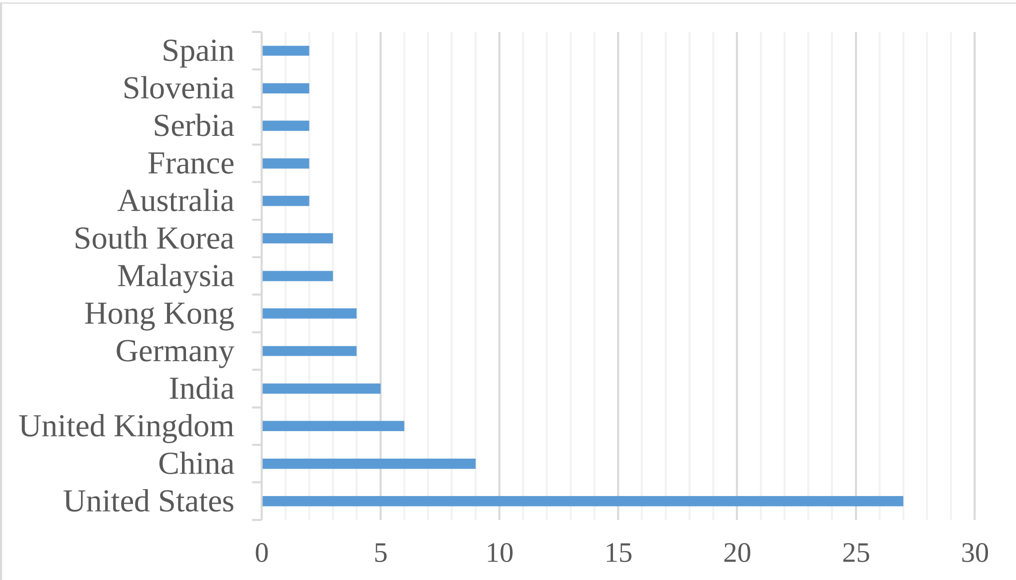


Figure 2.7 Classification of journal papers based on affiliation

2.3.2.4 Keyword statistics

Similarly, the keywords used in the selected journal papers were extracted using BibExcel, and the frequency of usage was analysed (Table 2.2). While the most frequently used keywords are ‘Big Data’ and ‘Supply chain management,’ it is evident that a wide range of other terms such as logistics, predictive analytics, RFID, and cloud computing were also used frequently.

Table 2.2 Keyword statistics

Keyword	Frequency	Keyword	Frequency
Big data	69	Logistics	10
Supply chain management	40	Internet of things	9
Supply chains	21	Radio Frequency Identification (RFID)	8
Manufacture	17	Big data analytics	7
Decision making	15	Supply chain	7
Data mining	14	Forecasting	7
Information management	12	Data handling	6
Data analytics	11	Cloud computing	6
Predictive analytics	11	Mass customization	4
Business analytics	10	Mass production	4

2.3.2.5 Theoretical Perspectives

The prominent theories used to explain the phenomenon of BDA adoption and practice were also investigated during the review process. Figure 2.8 shows that Resource-Based View (RBV) was extensively used by 38% of research papers. Besides that, Dynamic capability theory (19%), Information processing view (15%), Contingency theory (8%), social capital theory (4%), and other theories were also applied frequently. Apart from that, frameworks such as Technology Organisation Environment (TOE) were also used to investigate organisations' BDA adoption behaviour (Chen et al. 2015). Thus, there is much scope for application and validation of several other theoretical lenses such as Knowledge based-view, absorptive capacity, systems theory, institutional isomorphism, and agency theory to explore the phenomenon of BDA practice. Moreover, Hazen et al. (2016) have provided a review of several theories that can be applied in this domain.

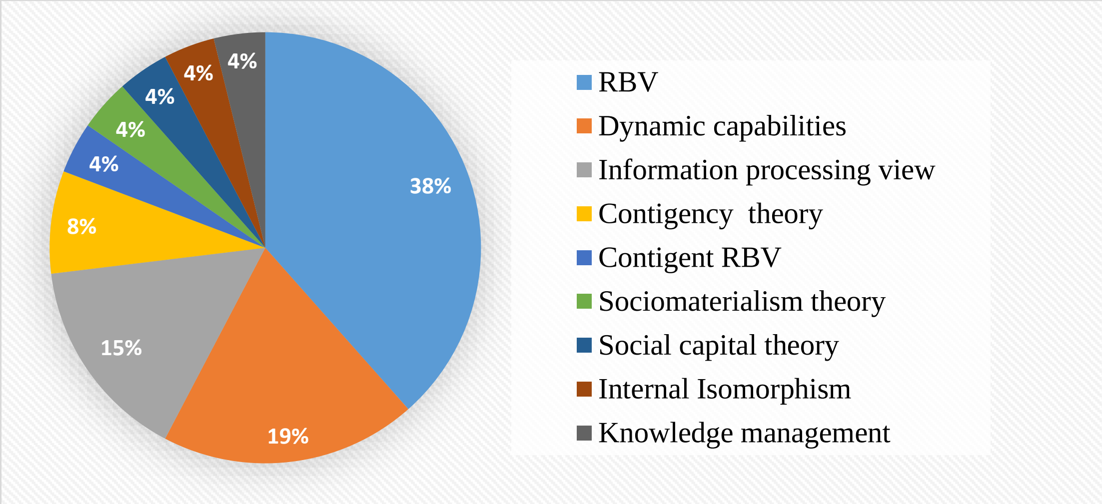


Figure 2.8 Frequency of underpinning theories in selected papers

2.3.2.6 Types of research methods

The research methods used by the selected papers were also reviewed. Few papers have used experimental, analytical or a mixed research approach. Interestingly, social media research has emerged as a new form of an academic research method to address supply chain issues. In particular, recent studies by Bhattacharjya et al. (2016), Mishra and Singh (2016), Papadopoulos et al.(2016), Chae (2015) and Chan et al. (2015) have focused on the use of social media (i.e. Twitter and Facebook) data and text mining approaches to address issues related to food supply chains, customer service and operations management. Likewise, case studies of integrating customer reviews data from commercial websites and transactional data for better demand prediction is also observed in the literature (Li et al. 2016, See-To and Ngai 2016). This change in research methods exhibits the influence of BDA on the academic community as well. However, the findings suggest that the majority of studies have used conceptual approach, followed by case study and Survey-based research (see Figure 2.9). Consequently, the current study falls into the category of survey-based research approach (Chapter 4-section 4.6).

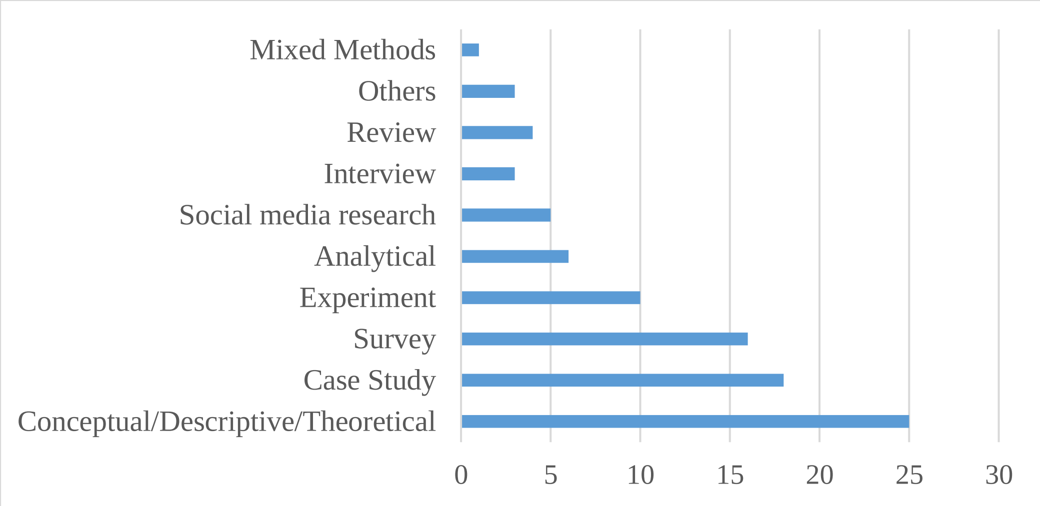


Figure 2.9 Classification of selected papers based on research methods

Moreover, among the survey-based studies, the distribution of papers based on the geographical representation of sample population was observed. Most of the studies have focused on global sample (n=4), followed by Iran (n=3) and other countries such as China (n=1), India (n=1), UK (n=1), Spain (n=1), and Germany (n=1). Several authors have recommended conducting studies in different geographical contexts and organisational cultures (Popovič et al. 2012; Bruque Cámara et al. 2015). Moreover, 54% of survey studies have centred on the manufacturing industry and the remaining 46% of studies have focused on all types of sectors as research context.

2.3.2.7 Top cited papers

From the selected papers, Hu et.al. (2014), Trkman et al. (2010) and Waller and Fawcett (2013 b) have received the highest number of citations (Figure 2.10). Trkman et al. (2010) is the first to quantitatively investigate the impact of business analytics on supply chain performance, followed by several others such as Oliveira et al. (2012), Chae et al. (2014 a), Chae et al. (2014 b), Jamehshooran et al. (2015 a), Jamehshooran et at. (2015 b), Cao et al. (2015), Sangari and Razmi (2015), Chen et al. (2015) and Gunasekaran et al. (2017).

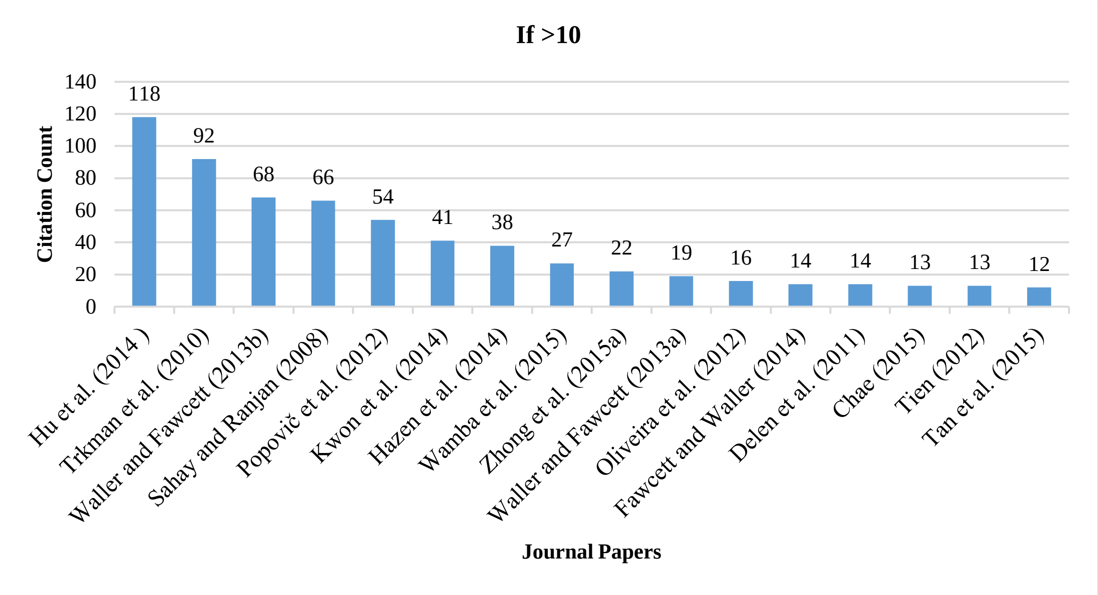


Figure 2.10 Distribution of papers based on the number of citations

Lastly, Figure 2.11 illustrates the contribution of journals based on the number of research papers. Among he selected papers, International Journal of Production Research (IJPR), Annals of Operations Research (AOR), International Journal of Production Economics (IJPE), Decision Support Systems (DSS), and Journal of Business Logistics (JBL) are the top 5 contributors in this domain. It was noticed that most of these journals have contributed to the domain mainly during recent years, indicating acceptance of BDA as a thriving research in operations research/management community. Moreover, considering the growing interests it is expected that the figures would rise in coming years.

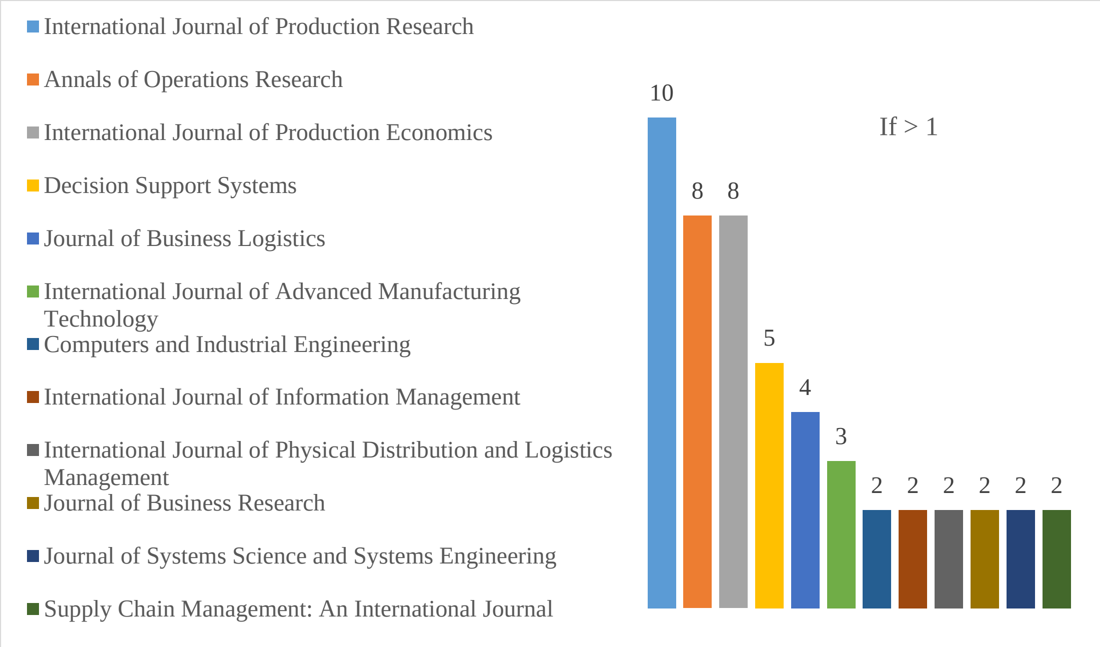


Figure 2.11 Contributions from different journals

2.3.3 Thematic Analysis

After the descriptive analysis, category selection and material evaluation are the third and fourth step in the review process towards the conceptualisation of the BDA concept. In this research, the thematic analysis was used to explore BDA practice to conceptualise BDA capabilities. The conceptual framework and its dimensions have evolved during the analysis by identifying themes and coding it along the review process. Since it is an emerging field, using a deductive approach would be inadequate because new codes have to be devised adaptably as they emerge from the data (Saunders et al. 2016). Therefore, in this research, a bottom-up/data-driven approach progressing from individual dimensions to the whole by reading and rereading the selected papers was used. The coding was verified independently by peer researchers to ensure reliability and validity (Seuring and Müller 2008). Minor amendments were made during the coding process to solve discrepancies between the researcher and the supervisors. For instance, some papers that are wrongly categorised into a theme were revised during the coding process in consultation with the supervisors. Finally, the papers were classified from a capabilities perspective into six key dimensions/themes; Data Generation (DG), Data Integration and Management (DIM), Advanced Analytics (AA), Data visualisation (DV), Data-Driven Culture (DDC), and Big Data Skills (BDS).

2.3.3.1 The conceptualisation of BDA Capabilities

This section presents the discussion around the capabilities identified from the thematic analysis. A conceptual framework was developed to simplify and delineate the BDA capabilities identified (Figure 2.12), and there are four quadrants in the framework; (i) Data poor and analytics poor, (ii) Data rich and analytics poor, (iii) Data poor and analytics rich, (iv) Data rich and analytics rich. The levels of BDA capabilities, measured by 5 dimensions, determine the quadrant within which organisations could find themselves. DG and DIM capabilities collectively constitute an organisation's level of data resources (data-poor or data-rich). Similarly, AA and DV capabilities denote an organisation's analytics resources (analytics poor or analytics rich). In Figure 2.12, data-driven culture is an intangible resource, fundamental to other capabilities, and has to be embedded into businesses routines related to BDA-driven decision making. Besides these, the cloud computing capability and Absorptive Capacity of organisations are considered as complementary resources and enablers of BDA. It can be argued that manufacturing organisations need to possess all key capabilities to obtain value or wisdom from raw data as illustrated in Figure 2.12.

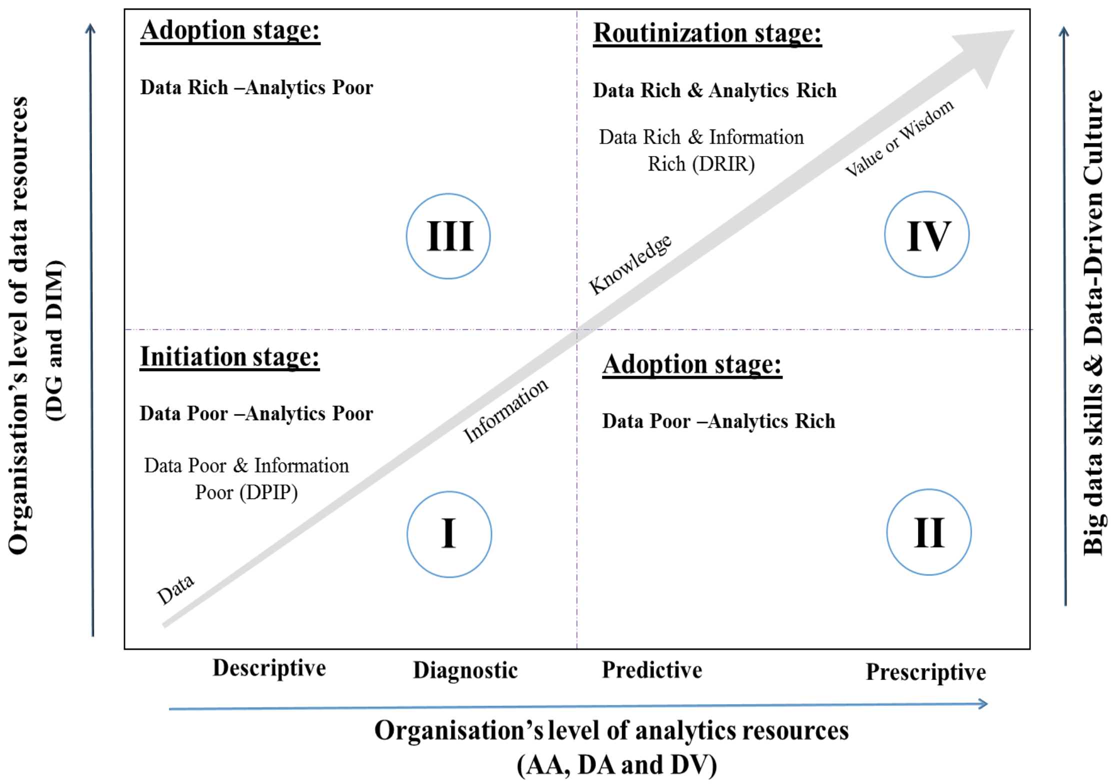


Figure 2.12 BDA capabilities framework

Table 2.3 Dimensions of BDA capabilities maturity

Dimensions of BDA capabilities maturity	Key elements	Key references
Data generation capability	1) Data generation infrastructures. 2) Data sources. 3) Repository for open data. 4) Strategies to collect data from sensors and other devices in real-time. 5) Data gathering capability.	Brandau and Tolujevs (2013); Hu et al.(2014); Radcliffe (2014); Janssen et al.(2014); Zhang et al.(2013); Jin and Ji (2013).
Data integration and management capabilities	1) Data integration from heterogeneous sources. 2) Data warehouse capability 3) Data management resources (capture, manipulation and redistribution of data) 4) Data governance 5) Unstructured data management 6) Data quality (accessibility, completeness, timeliness, reliability, consistency, and accuracy) 7) Metadata repository	Popovič et al.(2012); Spruit and Sacu (2015); Lavalle et al.(2010); Halper and Krishnan (2014); Radcliffe (2014); Sulaiman et al. (2015); Cosic et al. (2015); Nott (2014); IDC (2013); Knowledgent (2014); Tan et al.(2015); Chae et al. (2014a); Chae et al.(2014b)
Advanced analytics capabilities	1) Predictive analytics. 2) Real-time analytics. 3) In-memory analytics 4) Data mining (time series analysis, association rule mining, classification and clustering analysis). 5) Web mining and text mining. 6) Online Analytical Processing (OLAP). 7) Trend analysis and “What-if” scenario analysis. 8) Data analysis and Data decision capability.	Trkman et al. (2012); Oliveira et al. (2012); Wang et al.(2016a); Nott (2014); Radcliffe (2014); Knowledgent(2014); Sulaiman et al.(2015); Lavalle et al.(2010); Blackburn et al.(2015); van der Spoel et al.(2015); Zhang et al. (2013); Brandau and Tolujevs (2013); Popovič et al.(2012); Hu et al.(2014)
Data visualisation capabilities	1) Interactive visualisation. 2) Dashboards and key performance indicators (KPI). 3) Real-time information monitoring. 4) Strategic and operational reporting using historical and streaming data.	Brandau and Tolujevs(2013); Dutta and Bose (2015); Neaga et al. (2015); Tien(2012); Sulaiman et al.(2015); See-To and Ngai(2016); Radcliffe (2014); Popovič et al. (2012); Yesudas et al.(2014)
Big Data Skills	1) BDA Technical skills 2) BDA managérial skills	Sangari and Razmi (2015); (Gupta and George 2016)
Data-driven culture	1) Cultural and political issues. 2) Culture and execution. 3) Culture Capability.	Cosic et al.(2015); Dutta and Bose (2015); Halper and Krishnan (2014); Nott (2014); IDC (2013)

	4) Cultural competence in the context of supply chain.	

2.3.3.1.1 Data generation capability

Data Generation (DG) capability is the ability of organisations to seek, identify, create, and access data from heterogeneous data sources across organisational boundaries. Only if there is enough data, insights can be drawn from it. DG capability facilitates the availability of BD at organisations disposal by establishing data sources, procedures and policies to generate required data for decision-making. In general, data can come from three different domains ‘business, internet, and scientific research’ (Hu et al. 2014, p.657). Adoption of technologies such as Advanced planning and scheduling (APS), RFID, ERP, CRM systems and Warehouse management systems (WMS) (Autry et al. 2010), are the primary sources and antecedents for the occurrence of data deluge. Generation of data is further revolutionised with the advent of the Internet of Things (IoT) technology facilitating real-time sensing and transfer of events data.

Bauer et al (1994) argued that two types of data are generated in a manufacturing environment; static and dynamic data. The static manufacturing data contains information on manufacturing and assembling of products based on process times, product routings and structures, etc. Data related to ‘Bills of Material (BOM)’, ‘Bills of Process’, and ‘the factory level schedule’ are examples of static data generated in the production environment. Whereas examples of dynamic manufacturing data include ‘location of each work in progress batch’, and ‘the status of the input buffer and inter-cell transformation systems’. The dynamic data presents the current state of the manufacturing shop floor with information on quality, work-in-progress and performance measures. Figure 2.13 illustrates different types of data that can be generated in the manufacturing environment.

Moreover, in a manufacturing environment, data is scattered and can be acquired from diverse sources. The primary sources of data are from Enterprise information systems (EIS), which are mostly structured and transactional in nature. However, IoT, sensors, and RFID devices have the ability to convert the physical world into a virtual environment, which in turn generate a huge volume of unstructured data. Installation of RFID tags and readers on logistic objects can convert them into ‘passive smart logistics objects’ and ‘active smart logistics objects’ (Zhong et al. 2015). When these physical objects start communicating via wireless communication, an enormous volume of data gets generated in real-time making it difficult to handle. For instance, a manufacturing plant that has deployed 1,000 RFID readers and 10,000 tags could potentially generate Terabytes of data from a single day of operation (Zhong et al. 2015). It is reported that, by 2017, 50% of analytics implementation will utilise event-based streaming data (velocity) generated from machines or any physical objects that can communicate (Gartner 2014). The data generated by event logs are also used by manufacturing industries to perform process mining.

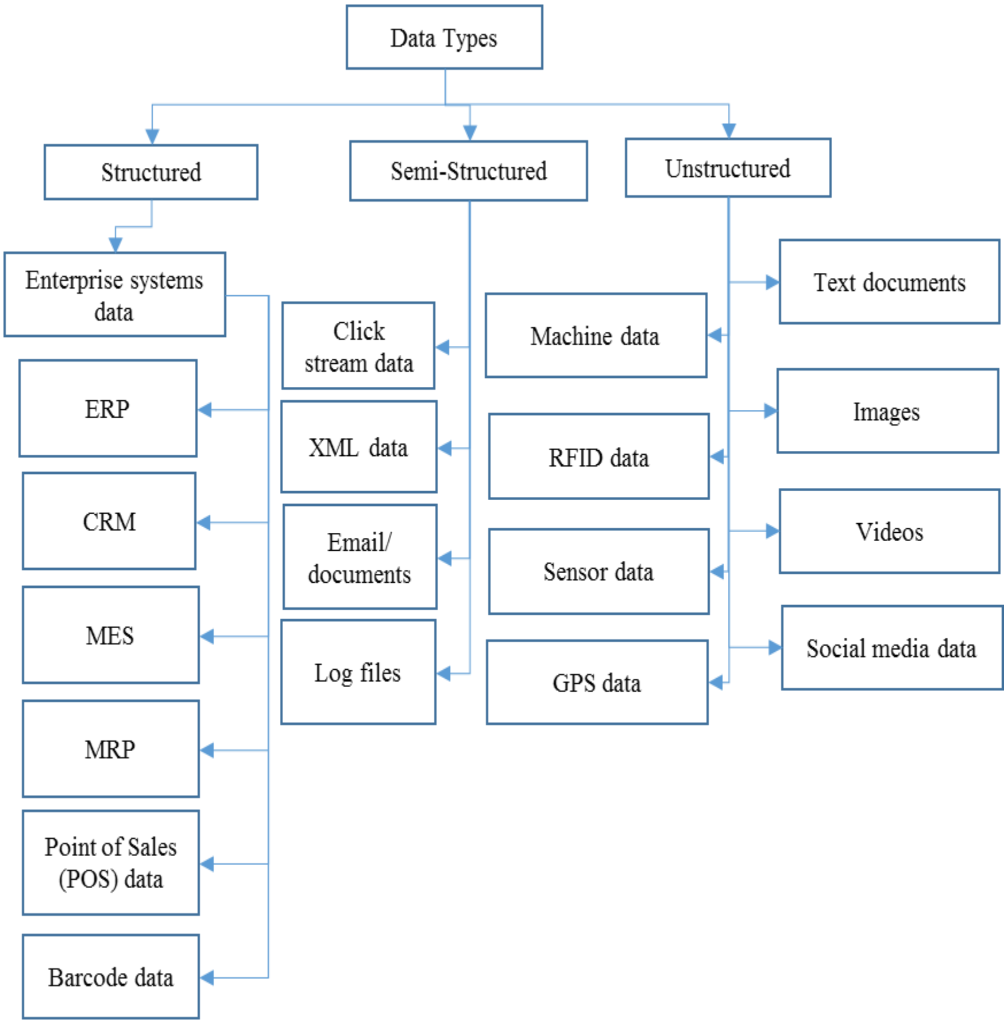


Figure 2.13 Different types of data generated in the manufacturing environment

Data is the raw material for the analytics system and the data generation capabilities determine the position of an organisation on the continuum of data poor to data rich. Organisations that have leveraged data generation capabilities are supposedly in a position to gain a competitive advantage by means of meticulously sensing their business environment, collecting data and information and exploiting it for business growth. Organisations should constantly seek, identify and access data from heterogeneous data sources. Data generation capability indicates a set of tools, technologies, and practices, which facilitate the generation of huge volumes, varieties, and velocities of data. In order to generate the right data for decision-making, organisations should put in place quality infrastructure and organisational practices to leverage the full potential of BDA. Based on the framework, it is argued that an organisation, which does not have the right infrastructure to support data generation and analytics, would fall into the quadrant of 'Data Poor and Information Poor'.

2.3.3.1.2 Data Integration and Management (DIM) capability

DIM capability is the ability of organisations to utilize tools and techniques to collect, integrate, transform and store data from heterogeneous data sources. The level of data integration, and ability to integrate different types of data gathered across organisational boundaries in real-time constitutes the DIM capabilities.

Vast amounts of data are distributed in heterogeneous sources and the integration of these isolated data will be challenging (Stefanovic 2014). DIM can be sub-divided into data acquisition from data sources, data pre-processing, and data storage. Due to the dynamic nature of the manufacturing environment, real-time access and scalability of data storage are the key capabilities to possess. Because of the complex nature of BD, the traditional database systems such as Relational Database Management Systems (RDBMS) are incompatible (Ge and Jackson 2014). Chae and Olson (2013) considered that Inter-organisational systems (IOS) such as Web-based or cloud-based EDI (Electronic Data Interchange) could be used to enhance data integration capability. Using processes such as Extract Transform Load (ETL) or Extract Load Transform (ELT) data can be migrated into OLAP/OLTP systems. Data integration capability can improve visibility, responsiveness, and performance of material management, and provide a 360-degree view of manufacturing operations (Xiong et al. 2015). Wamba et al. (2015) have elaborated via a case study that service delivery can be improved by integrating Intra- and Inter-organisational data. Tan et al. (2015) have exhibited the benefits of innovating products by integrating data from multiple sources such as internal consumer data, social media data, and multimedia data. Walmart integrates millions of transaction data generated every hour into one single system (Tee et al. 2007). Moreover, to process data in real-time, in-memory databases and distributed or parallel processing approaches are more appropriate than traditional approaches (Hahn and Packowski

2015). Data storage components can include a set of hardware and software infrastructure that is suitable for scalability, storing data of different types that aggregate batch wise, real-time or in near real-time, and also systems to support fast querying and retrieval of data (Hu et al. 2014). Advanced big data technologies and cloud computing based data storage, distribution, and retrieval of information can enable integration of data from various entities. Consequently, data acquired from heterogeneous systems has to be transformed into a standard format for further analysis. Data pre-processing activities such as ETL and ELT can be used to alter messy raw data into quality data sets suitable for analytics operations. However, the data integration capability has to deal with two main issues: data inaccuracy and redundancy (Chae and Olson 2013). BDA vendors suggest organisations must put emphasis on data governance too as it is the biggest challenge that an organisation could face (Meredith et al. 2012).

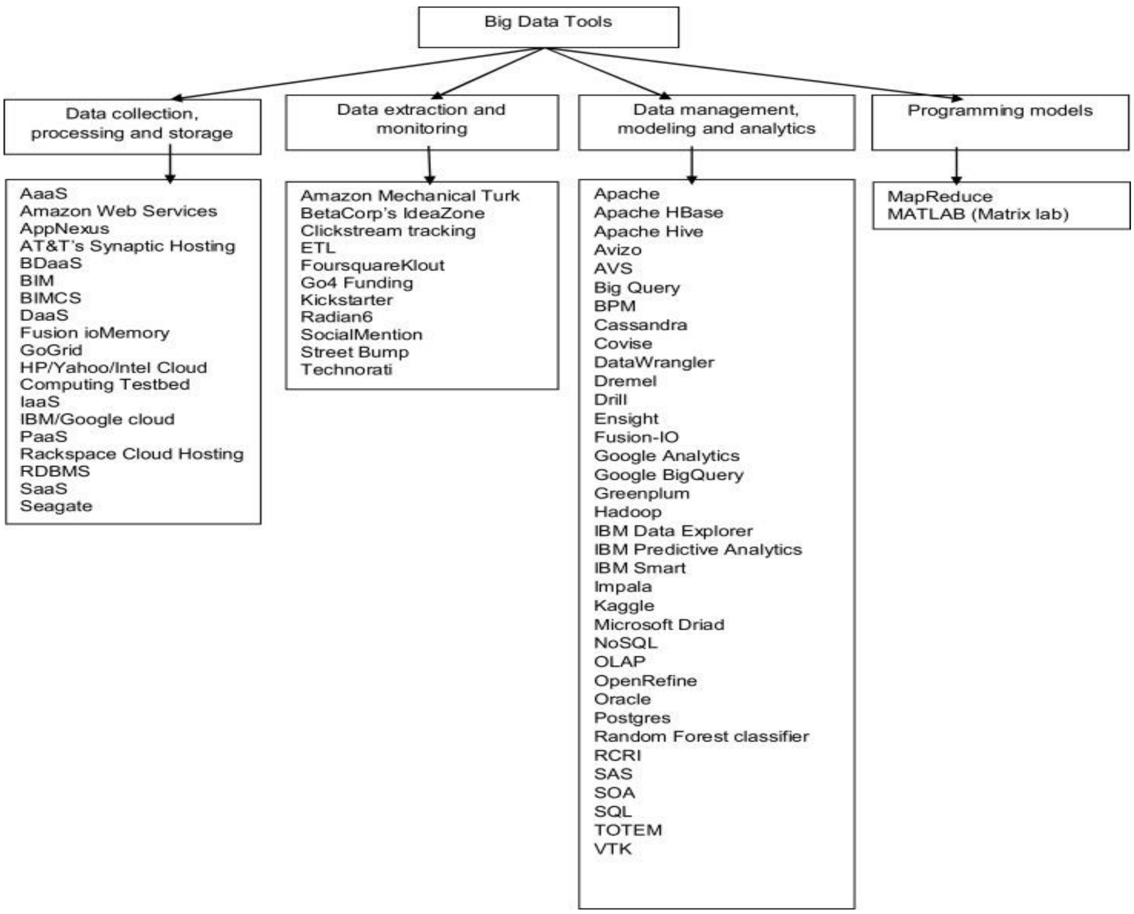


Figure 2.14 Big Data tools

Source: (Frizzo-Barker et al. 2016)

2.3.3.1.3 Advanced Analytics capability

Analysing Big Data is a complex task and requires advanced analytics capabilities. *Advanced Analytics capability* is the ability of organisations to utilise tools and techniques to analyse manufacturing data in batch wise, real-time, near-time, or as it flows and extracts meaningful insights for decision making. Data analytics is the most significant phase in the data value chain from raw data to meaningful insights; analytical tools and techniques are leveraged to slice through the data to convert into data-driven insights. Depending on the depth of analysis, data analytics techniques are classified into descriptive, predictive and prescriptive analytics Table 2.4 (Souza 2014). As shown in Figure 2.12, with the increase in analytics capabilities from descriptive to prescriptive analytics, organisations' decision-making capabilities at operational (short-term), tactical (mid-term) and strategic (long-term) level would certainly improve. Descriptive analytics is based truly on the principle of classical statistics methods. However, predictive analytics is the combination of statistics, data mining, and machine learning techniques (Blackburn et al. 2015). While descriptive analytics rely on historical data, predictive analytics can utilise historical data and in addition, dynamically include external data from other sources but is contingent upon the analytics models and algorithms used. Manyika et al. (2011) discussed several analytics techniques available to analyse Big Data such as association rule mining, classification, clustering, crowdsourcing (data collected through web 2.0), data fusion and integration, ensemble learning, machine learning, genetic algorithm, network analysis, neural network, predictive modelling, sentiment analysis, spatial analysis, and time series analysis.

Predictive analytics capability enables organisations to consider both endogenous and exogenous variables while forecasting demand. The unstructured customer reviews have variables that can predict sales 'nowcasting' (See-To and Ngai 2016). Traditional forecasting depends on aggregated data, but by deploying real-time analytics capabilities organisations can analyse demand data in real-time increasing accuracy, potentially reducing the bullwhip effect (Hofmann 2015). Increasing the robustness of demand forecasting via predictive real-time analytics can eventually improve other functions such as production planning and inventory optimisation which rely on the forecast demand. Moreover, since a huge volume of spatiotemporal data is generated from GPS and RFID devices, predictive and spatiotemporal analytics can be used to analyse these unique data types, for instance, to predict truck arrival times (van der Spoel et al. 2015) and optimal sourcing of blood (Delen et al. 2011). Information access and content quality are empirically found to increase by leveraging data integration and analytics capabilities (Popovič et al. 2012). It is also found to increase an organisation’s information processing capabilities (Cao et al. 2015) and manufacturing planning satisfaction (Chae, Yang and Olson 2014).

Table 2.4 Different kinds of analytics

Kinds of Analytics	Purpose
Descriptive analytics	Refers to a descriptive summary of events or situations that lie in the past
Predictive analytics	Future-oriented attempting to forecast how an event or a situation will look like
Prescriptive analytics	Employing more complex algorithms, including machine learning, to under-stand the past and advise which action out of multiple alternatives will yield the optimal outcome

Source: adapted from (Souza 2014)

In manufacturing organisations, machine learning algorithms can be used for system diagnostics, machine condition monitoring, classification, dimension reduction, state prediction and process planning (Chan et al. 2018). Deng and Yeh (2011) used an advanced least-square Support Vector Machine (SVM) algorithm to predict manufacturing costs. Similarly, techniques such as k-nearest neighbour are used to predict the cost of the welding process in the manufacturing context (Sajadfar and Ma 2015). Chien et al. (2015) utilised advanced analytics to detect the root cause of downtime to enhance the yield. Gunasekaran et al. (2018) have demonstrated that the BDA can enhance agile manufacturing practices to improve performance.

Tan et al. (2015) described how text analytics capability could be combined with other analytics capabilities to develop innovative products. Markham et al. (2015) also emphasised the significance of text analytics capabilities to make new product development decisions. It is also suggested that the integration of Twitter data mining capabilities into CRM systems would be beneficial to resolve queries in real-time and improve customer satisfaction (Bhattacharjya et al. 2016). Moreover, applying predictive analytics (decision tree algorithm) on unstructured data from commercial websites, which include transactional as well as behavioural data of consumers, can help e-commerce businesses to identify, predict and manage demand (Li et al. 2016). Boone et al. (2016) claimed that BDA could be beneficial to enhance the performance of after-sales service parts management. However, Jain et al. (2017) argued that most manufacturing organisations do not completely utilise all the data generated such as data from digital manufacturing systems and computer-aid manufacturing. The level of analytics capability determines the ability of the manufacturing firms to extract value from the data generated from various sources. Hence, as argued by Markham et al. (2015), data collection (Big data) is only the first step in the data value chain, and advanced analytics capabilities are required to extract value from it.

2.3.3.1.4 Data Visualisation capability

Data Visualisation capability is the ability of organisations to utilise tools and techniques to visually render information and deliver the data-driven insights intuitively in a timely manner to the decision makers. Data visualisation is "the representation and presentation of data that exploits our visual perception abilities in order to amplify cognition" (Kirk 2012, p.17). The two main purpose of data visualisation, which is a form of descriptive analytics, is sense-making (data analysis) and communication of abstract information through the graphical display (Few 2009).

In a manufacturing context, data visualisation capabilities are equally important as other BDA capabilities. They can be beneficial in various ways such as monitoring operations, visualise material movement and material tracking. The primary goal of data visualisation is to intuitively represent knowledge and information using various techniques such as tag cloud, graphs, Clustergram, and heat maps (Chen and Zhang 2014; Manyika et al. 2011). In particular, heat maps which could display geographical information such as consumption location, transaction density, etc. is valuable to develop new distribution strategies (Tachizawa et al. 2015). Tools such as desktop applications, interactive web and mobile applications offer capabilities to executives, decision makers and customers to interact with the analytics ecosystem (Barlow 2013). The presence of data visualisation capabilities can effectively support the three analytics capabilities (descriptive, predictive, and prescriptive) (Assunção et al. 2015). From the usability perspective, highly interactive visualisation would help users to interpret, find patterns and make decisions quickly and effectively from raw data as well as content from the analytics system.

Manyika et al. (2011) emphasised that insight should be presented in such a way that people can effectively consume it and aid them to take action. Data visualisation tools are significant to monitor key performance indicators (Xiong et al. 2015). eBay leveraged Tableau (a popular data visualisation platform) to transform large complex data sets into interesting pictures for decision-makers to understand consumer behaviour and trends via techniques such as sentiment analysis (Chen and Zhang 2014). Brandau and Tolujevs (2013) experimented with visualisation techniques and clustering algorithms to manage irregularities in real-time sensor data and improve logistics performance. Moreover, Bock and Isik (2015) have justified the importance of visualisation in informing managers about the current status of processes. The paper argues that the ability to create and transfer the visual representation of analytics insights in real-time reduces the time difference between the process of data collection, storage, analytics, and reporting. For SCM, integrating the large volume of data and visualising actionable insights should be the priority to develop operational intelligence solutions (Yesudas et al. 2014). In fact, it is widely accepted that many organisations prefer to use intuitive data visualisation tools (such as Tableau, Spotfire, QlikView, Datameer, etc.) that are catered to the needs of information visualisation problems (Zhong et al. 2016 b).

2.3.3.1.5 Data-Driven Culture

Data-driven culture is an intangible resource that represents the beliefs, attitudes, and opinion of people towards data-driven decision-making. According to Aho (2015, p.284) "The transformative potential of Big Data lies in treating data as an asset." Aho (2015) argued that Big Data involves extensive change management and development of a new organisational culture to transform the organisation. According to Kiron and Shockley (2011), three main elements that constitute to data-driven culture, namely: (i) Analytics should be treated as a valuable or strategic assets by an organisation, (ii) Top Management should support and leverage analytics across the organisation, and (iii) Accessibility of Data-Driven insights to decision makers. Real-world case examples have suggested engagement of the implementation team and top management support are significant for developing BDA capabilities (Wamba et al. 2015). Lavalle et al. (2010) conducted a survey to identify the significant barriers to the adoption of analytics in organisations. Despite the challenges of technology and data quality, organisations considered culture and managerial issues as the significant barrier to analytics adoption. To build transformational BDA capabilities, business and IT leaders in organisations have to work together and develop new strategies and roles such as Chief Data Officer, and Data Scientists to address the needs of technology and business. Moreover, Cao et al. (2015) argued that the presence of a data-driven culture would facilitate organisations to make data-driven decisions and rely on fact-based decisions to develop new products and services. Certainly, organisations who possess advanced analytics capabilities could not extract full value if they are not effectively integrated into the business decision-making process and are not accepted as a decision-making tool (Blackburn et al. 2015). It is imperative not only to develop a data-driven culture within an organisation but also across organisational boundaries. For example, Walmart extends their analytics capabilities to all of its suppliers and promotes the culture of data-driven decision-making to improve performance.

2.3.3.1.6 Big Data Skills

Human capital sets the foundation for organisations to attain competitive advantage (Liu 2014). Experienced personnel who possess skills, knowledge and capabilities can create economic value for organisations via improved productivity (Youndt et al. 1996). Becker (1975) has discussed two categories of human capital; general and specific human capital. General human capital is transferrable across organisations, but specific human capital is non-transferable and related to firm-specific experiences. Although the adoption and practice of BDA inherently involve many challenges, employees must address those challenges by harnessing their skills. Without the knowledge input of skilled personnel, the investment made to create BDA resources would fail to produce economic value. Further, intangible tasks performed by the skilled personnel such as problem-solving, and critical judgment form a major portion of large firms' everyday operations (Wang et al. 2014). Especially, the role of skilled personnel is significant in addressing uncertainties and it requires personnel to adapt to rapid change in technology and strategy (Shen et al. 2010).

Wamba et al. (2017) considered BDA personnel capability as one of the dimensions of BDA and referred to it as "the BDA staff's professional ability (e.g., skills or knowledge) to undertake assigned tasks" (p.3). Moreover, some of the BDA maturity models reviewed in this study considered skills as a one the main dimensions of BDA. For instance, (Halper and Krishnan 2014)'s maturity model include people as a main component of BDA maturity along with process and object. Similarly, in maturity models developed by Howson (2015), IDC (2013), Radcliffe (2014) and Cosic et al. (2015), people skill is considered as an important element when assessing the BDA maturity level of an organisation. In terms of types of skills, studies on IT literature has categorised skills into two major types: 1. Technical and 2. Managerial skills (Pavlou and El Sawy 2006). Similarly, some literature on BDA has investigated the role of both technical and managerial skills of BDA (Gupta and George 2016). Tomar et al. (2017) summarised the results of a few studies that investigated the influence of human elements such as skills and knowledge on extracting value from BDA. It is found out that technical, analytical and governance skills are much-needed skills to reveal the full potential of BDA. Other studies have also suggested that BDA personnel skills are associated with the competitive advantage gained from the exploitation of BDA resources (Sangari and Razmi 2015; Gupta and George 2016).

In general, the creation of data products needs efforts from various professionals within an organisation. This may include involvement from programmers, domain experts, and data scientists. Programmers help in activities such as data assimilation, integration and consumption. Domain experts help in authenticating the data generation and merging data from various locations without losing the contextual information. On the other hand, data scientists apply various analytical techniques such as predictive modelling on the data and perform various related activities such as validating predictions, dealing with missing data, etc. Waller and Fawcett (2013b) argued that a data scientist requires a combination of both analytical skills and domain knowledge, which is difficult to find as someone good in analytical skills may not be interested in learning domain knowledge. Moreover, Schoenherr and Speier-Pero (2015) have empirically found that data scientists should have the following skill set: 1. Understanding the application of qualitative and quantitative methods of forecasting, 2. Numerical methods of optimization, 3. Broad awareness of many different methods of estimation and sampling, 4. Determining opportunity cost, 5. Using numerical methods to estimate functions relating independent variables to dependent variables, 6. Using probability theory with actual data to estimate the expected value of random variables of interest, 7.Quick design and implementation of discrete event simulation models, 8. Capital budgeting, 9. Managerial accounting, 10. Marketing science. So, typically, the people carrying out these roles come from several backgrounds with skills ranging from SQL, Python, R, Java, Scala, Hadoop, and so on (Tomar et al. 2017). Hence, BDA skills are also considered in this research as an important element of BDA maturity.

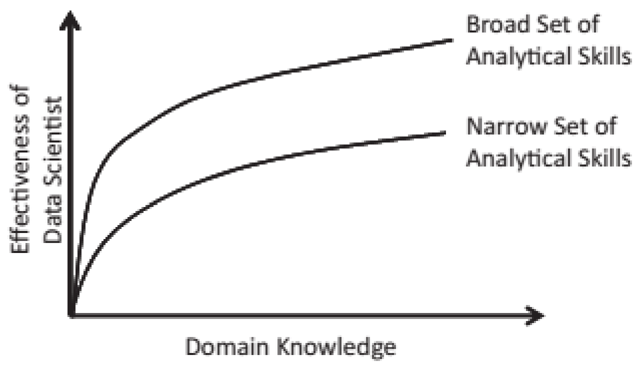


Figure 2.15 Data scientists effectiveness and domain knowledge

Source: (Waller and Fawcett 2013a)

2.4. Data and information quality

Data and information quality is truly a unique asset possessed by a company. While a product of a company is regularly mimicked by its competitors, the services of the company can be distinguished contingent upon the availability of good quality data. In general, a high degree of reliability and confidence in your data is required as it has a high impact on operational capabilities (Hubley 2001).

In the literature, there are several perspectives regarding the definition of data quality such as the quality of raw data and data products or information (Lee 2003). According to Batini and Scannapieco (2006, p.6), “data quality is a multifaceted concept, as in whose definition different dimensions concur”. Consequently, there is a wide range of definitions available in the literature. However, as this research is related to computerised database systems, where data quality control is of greatest importance, the definition accepted by the database community is considered as the appropriate one for this study. Redman (2008, p.4) defined data quality from the perspective of fitness-for use as follows: “Exactly the right data and information in exactly the right place at the right time and in the right format to complete an operation, serve a customer, make a decision, or set and execute strategy”. In the context of information systems and database engineering, Kenett and Shmueli (2017) stated that data quality refers to “the usefulness of queried data to the person querying it” (p.21). In the database community, there is no clear distinction between data quality and information quality, and the review of the literature reveals that the dimensions of both these concepts are quite similar. Moreover, it can be argued that, in practice, the terms ‘data’ and ‘information’ are used interchangeably. As shown in Figure 2.16, the quality of processed data products and information are used interchangeably in the literature, while these are certainly different from the raw data. Hence, this research views it as a single concept with insignificant variations and focuses on examining the role of data and information quality. Unless specifically addressed, any mention of ‘data quality’ will refer to the quality of data products and information.

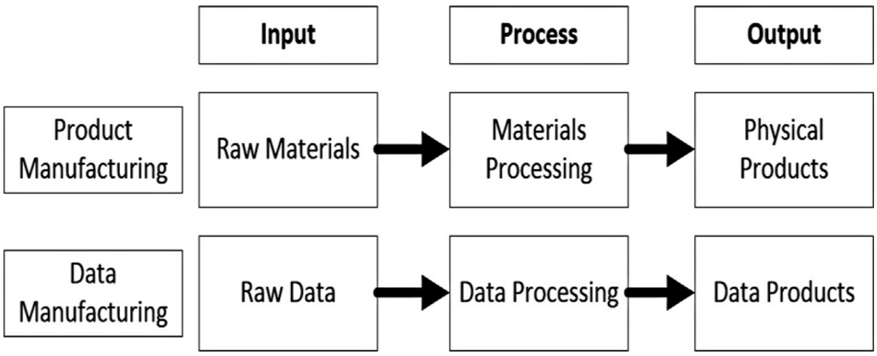


Figure 2.16 Analogy between the process of product and data manufacturing

Source: (Hazen et al. 2014)

Hazen et al. (2014) have argued that there are various similarities between the manufacturing of products and data. But, two main differences exist between these processes. In product manufacturing, the raw material is the input, whereas in the data production process raw data is the input. Further, measuring the quality of tangible manufactured goods is relatively uncomplicated compared to measuring the quality of data products which are intangible in nature. The efficiency of the physical flow of material is mainly determined by the infrastructure quality (such as transportation system, ports, technology, etc.) (Bagchi et al. 2014). Similarly, the efficiency of information flow demands quality data infrastructures such as NoSQL, RDBMS, and Hadoop. Different types of data based on the data lifecycle approach is presented in Table 2.5.

Table 2.5 Types of data based on data lifecycle

Types of Data	Definition
Raw data items	Smaller data units which are used to create information and component data items
Component data items	Data is constructed from raw data items and stored temporarily until the final product is manufactured
Information products	Data, which is the consequence of performing manufacturing activity on data

Source: (Sidi et al. 2012)

The system model given in Figure 2.17 indicates the process of transforming raw data into insights. This process may involve various activities of transformation. Data cleaning and pre-processing, storage, and analytics are sub-processes within the data operational process, and performance should be monitored at every stage to maintain the quality and reliability of resultant data and information products. Whereas in the case of data streams and real-time analytics (continuous process), the data generation process (which precede the transformation process) has to be monitored.

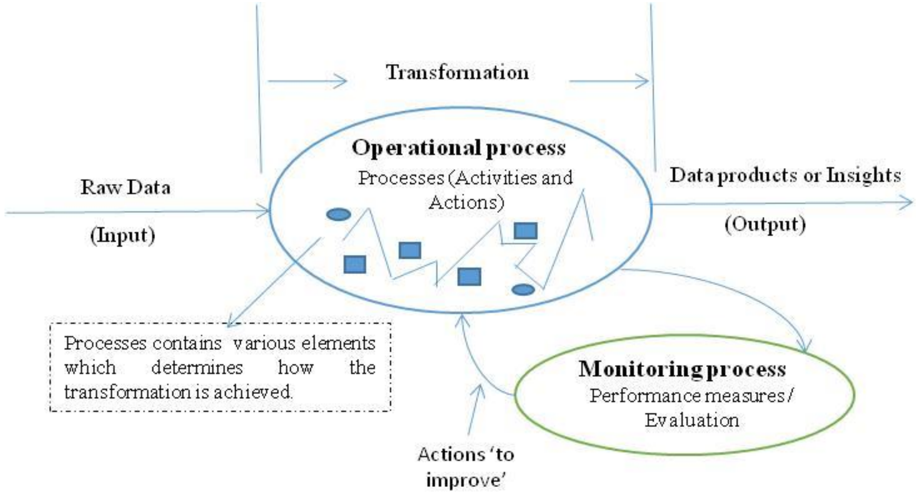


Figure 2.17 A systems approach of the data transformation process

Source: adapted from (Kawalek 2008)

From the literature, it is evident that both the dimensions of data and information quality are identical. As shown in Table 2.6, scholars have identified various dimensions of data and information quality. Wang and Strong (1996) and Lee et al. (2002) have classified the dimensions of DIQ into *intrinsic* (accuracy, timeliness, consistency, and completeness) and *contextual* (relevancy, value-added, quantity, believe-ability, accessibility, and reputation of the data) dimensions. Measurement of c**ontextual dimensions** relies on subjective measures via self-report surveys, user questionnaires, and situational judgement of decision makers. Whereas, the measurement of **intrinsic dimensions** relies on various monitoring tools like total quality approaches, process capability analyses, and statistical process control (SPC). Similarly, Kenett and Shmueli (2017) mentioned that the dimensions of information quality can be categorised into intrinsic, contextual, representational and accessibility. Moreover, information quality indicates that "the user consider the data in the context of the user, rather than in isolation (as the term data quality might imply)" (Kenett and Shmueli 2017, p.22). However, there are five main dimensions such as accuracy, reliability, timeliness, completeness, and consistency frequently used by researchers to measure data and information quality (Wand and Wang 1996). According to Wand and Wang (1996), ***accuracy*** implies that "the information system does not represent a real-world state different from the one that should have been represented. (p.93)" It is related to correctness and focuses on precision to a reality of interest. Inaccuracy refers to mapping real-world systems incorrectly on to information systems. ***Reliability*** indicates "whether the data can be counted on to convey the right information; it can be viewed as correctness of data (p.93)." It is related to the capability of avoiding errors and indicates whether data can be counted on for further analysis. ***Timeliness*** refers to "the delay between a change of the real-world state and the resulting modification of the information system state (p.93)." Timeliness is affected by the volatility of the real-world systems, time of data usage and whether the information system is up-to-date. If there is a lack of timeliness, a past real-world state is reflected on the information systems state. ***Completeness*** is "the ability of an information system to represent every meaningful state of the represented real-world system (p.93)." In the manufacturing context, completeness is defined as "The extent to which the information is comprehensive for the planning tasks" (Gustavsson and Wänström 2009, p.331). Also, it represents the capability of capturing comprehensively all the relevant aspects of the reality. Completeness can be viewed from different perspectives such as schema completeness, column completeness, and population completeness (Lee et al. 2006; Batini and Scannapieco 2006). ***Consistency*** of data values means the mapping of the real-world state to only one matching information system state not one-to-many. Consistency represents the "extent to which data is presented in the same format and compatible with previous data" (Wang and Strong 1996, p.93)

Table 2.6 Dimensions of data and information quality

Category	Dimension	Definition: the extent to which …
Intrinsic	Believability	data are accepted or regarded as true, real and credible data
	Accuracy	data are correct, reliable and certified free of error
	Objectivity	data are unbiased and impartial
	Reputation	data are trusted or highly regarded in terms of their source and content
Contextual	Value-added	data are beneficial and provide advantages for their use data
	Relevancy	data are applicable and useful for the task at hand
	Timeliness	the age of the data is appropriate for the task at hand
	Completeness	data are of sufficient depth, breadth, and scope for the task at hand
	Appropriate amount of data	the quantity or volume of available data is appropriate
Representational	Interpretability	data are in appropriate language and unit and the data definitions are clear

	Ease of understanding	data are clear without ambiguity and easily comprehended
	Representational consistency	data are always presented in the same format and are compatible with the previous data
	Concise representation	data are compactly represented without being overwhelmed
Accessibility	Accessibility	data are available or easily and quickly retrieved
	Access security	access to data can be restricted and hence kept secure

Source: (Batini and Scannapieco 2016, p.40)

Moreover, scholars have recognised the importance of maintaining data and information quality. Redman (1998) has elaborated on the impact of poor data quality on various levels such as the operational, tactical, and strategic levels. The authors argued that at the operational level, a lack of data quality leads to various issues such as customer dissatisfaction, employee dissatisfaction, and increased cost. At the tactical level, poor data quality affects the decision-making process, makes it difficult to carry out re-engineering projects, and increases mistrust between various departments of an organisation. It is argued that managers spend half of their decision-making time arguing about the quality of data, affecting the effectiveness of the decision-making process. In a manufacturing facility, different departments rely on data from each other. For instance, take two departments A and B, needing data on specific parts that are being manufactured. Each department needs information on specific variables that the other department does not need. Department B may be disappointed with the quality of data maintained by department A and this increases the mistrust between them. At the strategic level, there is less evidence of the direct impact of poor data quality. However, the collective influence of issues created at the operational and tactical level due to poor quality can have an adverse effect on decision-making at the strategic level. A report from the Data Warehousing Institute suggests that US businesses incur an estimated cost of 600 billion US dollars due to data quality problems (Batini and Scannapieco 2006). Hazen et al. (2014) argued that data quality problems could hinder the data analytics activities and affect management decisions. The authors have compared the similarity between the manufacturing process and the data production process; raw materials are the input in the manufacturing process and, in the data production process, data is the input. However, it can be argued that the raw data is often a by-product of IT systems (ERP, CRM, etc.,) and DPB activities rely on these data resources. Nevertheless, apart from these endogenous data, DPB activities are capable of exploiting exogenous data available outside the organisation (Open data, Social media data, etc.). Unlike a physical product, data is intangible in nature, measuring the data quality is a multidimensional problem.

For experimental purpose, Hazen et al. (2014), conducted a study in the context of jet engines and components remanufacturing and tracked the data generation process focusing on engine location and repair status information. Data of eight jet engine compressors are gathered in real time and one of the intrinsic dimensions, completeness, was measured. Using the control charts (Bernoulli CUSUM), the incomplete records are monitored. By using this technique, the managers are able to identify the data quality problem that occurred while monitoring the observations of 6 jet engine compressors (control chart of jet engine compressors), which allowed them to take counteractive actions.

Hence, the assessment of data quality necessitates evaluation of various dimensions of it. Among the various dimensions discussed earlier, each organisation should determine which is more important for its operations and define the variables that represent the chosen dimensions. Only the finalised variables are measured as the concept is multidimensional in nature. It is argued that choosing which variables to measure is a complex issue, highly context-dependent and varies from organisation to organisation. Haryadi et al. (2016) grouped the antecedents of data quality into 5 categories: 1. data, 2) technology, 3) people, 4) organisation, and 5) external environment. Redman (2007) argued that technology can improve data quality. For instance, technological advances in the electronic recording of data, RFID, and data verification technologies have the potential to produce significantly ‘cleaner’ data than traditional manual data entry methods.

2.5 Supply chain analytics / Data-driven supply chain capabilities

In recent times, Big Data Analytics has been transforming various business functions, and Supply Chain Management (SCM) is one of the functions that is being transformed (Sanders 2014). In an ever-increasing competitive marketplace, organisations have clearly recognised the importance of SCM as a motivating factor to gain competitive advantage (Jamehshooran et al. 2015b). Achieving a competitive advantage at the supply chain level cannot be possible if it is devoid of data-driven supply chain capabilities. A supply chain is a complex network of organisations and resources that are aligned to act upon a cohesive process of transforming raw materials into products and effectively delivering it to customers. In a typical supply chain, there is a flow of information along with material and financial flows (Souza 2014). Supply chain analytics is the use of information and analytical tools to support efficient flow of material along the supply chain. Organisational boundaries cause fragmentation of data and processes (Li et al. 2015). According to Li et al. (2015, p.24), “The topology of supply chains has become a complex network with multiple starting points, multiple ending points, and numerous routines from start to finish. At the same time, the topology of supply chain networks has become more dynamic. Partner companies join and quit the network dynamically. Snapshots of the network change frequently.”

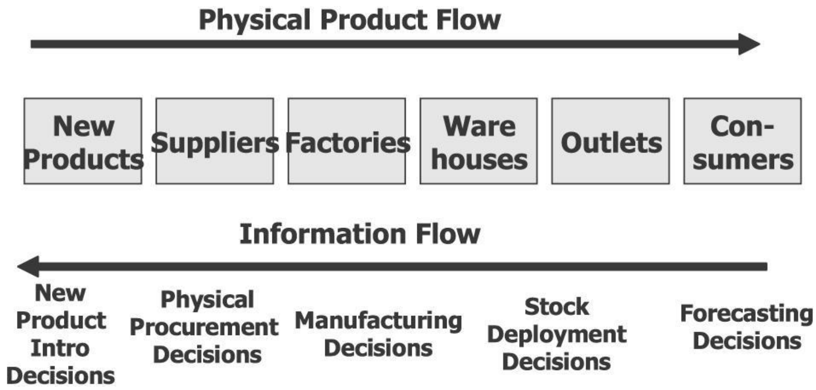


Figure 2.18 Physical flow vs information flow in a Supply chain

The adoption of electronic supply chain management (e-SCM) such as Internet-based inter-organisational systems and Internet-based electronic data interchange (EDI), has enhanced communication, coordination, and collaboration across organisational boundaries (Lin 2014). The adoption of these systems also generates a large volume of data through data exchange from the members of the supply chain network. Supply chain practices such as collaborative planning, forecasting, and replenishment (CPFR) also generate additional data that needs to be stored and monitored (Chae and Olson 2013). Apart from these, the ability to collect data from the internet and its significance in improving supply chain performance has also been addressed recently (Mishra and Singh 2016; Bhattacharjya et al. 2016). In addition, sharing of data generated from an entity to other parties in a supply chain could have a profound effect on optimising supply chain performance. Radke and Tseng (2015) elaborated the benefits of cloud computing for the purpose of data sharing in a more secure and agile manner within a supply chain network. In the supply chain context, there could be several reasons for reluctance in sharing data among its members. Some of the major barriers to data sharing are sensitivity, fear of losing competitive advantage and problems with access and control of information sharing (Radke and Tseng 2015). Supply chains are dynamic in nature and prone to environmental uncertainty. So, to streamline the data sharing process, it is imperative to have data access control for firms in a short-term relationship and limit the access when the relationship perishes. If data is scattered and not integrated to provide a single point of truth, then organisations would lack a valuable resource for decision making, i.e. quality data; therefore, SCM demands DIM capabilities.

Chen et al. (2015) have conceptualised the use of BDA in SCM into three categories: (i) Coordination/Integration process (Warehouse operations improvements, Process/equipment monitoring, and logistics improvements), (ii) Learning processes (sourcing analysis, purchasing spend analytics, CRM/customer/patient analysis, forecasting/demand management - S&OP, and Inventory optimization), (iii) Reconfiguration processes (network design/optimization, production run optimization, inventory optimization). An extensive multistage study conducted by Sanders (2016) illustrates real-world BDA applications in various areas of SCM including, but not limited to, inventory optimisation, labour scheduling, route optimisation, price optimization and micro-segmentation in marketing. Moreover, the accuracy of demand forecasting, one of the critical aspects of SCM, can be improved by using advanced predictive analytics techniques that outperform historical data based statistical techniques (Blackburn et al. 2015).

Table 2.7 Analytics use cases in SCM context

Analytics use cases in SCM context	Data types	Analysis techniques	Purpose
Monitor-and-navigate	Location data extracted from GPS and RFID.	Descriptive Analytics: Continuous monitoring of performance metrics, root causes analysis.	Increases visibility of assets and material flow in the supply chain.

Sense-and-respond use (Reactive)	Integrates internal sales data, POS data, Market research, and social media data.	Predictive analytics: Data mining and modelling techniques which use Multivariate statistical methods.	To discover knowledge and patterns that can be translated into business rules for (semi-)automatic responses to predefined business events. Moreover, analytics applications (in particular in-memory analytics) can be used to tracks supply risks, operations risk, and demand risk.
Predict-and-act (Proactive)	Sales data, operational data, weather data.	Predictive analytics: Time series analysis, scenario and risk profile analysis, Monte Carlo, discrete-event simulation, Cluster and regression analysis (to develop pricing strategies using sales data), correlation analysis.	To apply business forecasting and simulation methods to support business decisions, especially sales and operations management. • It supports strategic procurement, • Predictive maintenance and assets utilization status • Effective delivery, productivity, quality, and costs. • Used for sales forecast and effective stock management and stock reallocation. • Demand forecasting.
Plan-and-optimize	1) Strategic planning uses aggregate data. 2) Operational planning uses granular data (such as transactional data) obtained from everyday business activities.	Prescriptive analytics: Mathematical optimization techniques, What-if analysis, scenario analysis, APS systems,	Used for strategic planning (to supports the configuration of supply chain networks) and operational planning (to support material management and financial flows along the network). • It supports interactive analyses and scenario simulation s, • Strengthens collaboration, • Improves traceability, and enhances cross-functional coordination.

Source: adapted and modified from (Hahn and Packowski 2015)

Park et al. (2016) have developed a visual analytics-based decision support system (DSS) that incorporates predictive analytics capabilities and have experimented with supply chain network data. They argued that interactive visualisation would enhance the level of human cognition during decision-making. Similarly, GIS-based analytics and visualisation capabilities are found to be beneficial to effectively managing blood supply chains (Delen et al. 2011). Zhang et al. (2013) used data visualisation techniques to identify sources of contamination in the food supply chain. Due to dynamic nature, SCM demands real-time data analytics capabilities, which enables processing of data such as RFID data to monitor processes and events (Manyika et al. 2011). As given in Table 2.7, Hahn and Packowski (2015) discussed the analytics application for real-time data processing in a supply chain context, and categorised its application into: (i) Plan and optimise (ii) Predict and act, (iii) Sense and respond, (iv) Monitor and navigate. Accordingly, analytics on RFID and GPS data would enable continuous monitoring of material flow in supply chain, increasing the visibility of assets. Analytics can help organisations to sense and react to supply risk, demand risk, and operations risk, and can also be used to predict delivery, asset utilisation, maintenance, productivity, and forecast sales. For instance, data mining techniques can be applied in the area of supply chain fraud detection to reduce supply chain risks (Kraus and Valverde 2014). Similarly, leveraging prescriptive analytics can be beneficial for planning and optimising supply chains either at an operational or a strategic level using aggregate data from transactions and granular data from day to day business operations.

Similarly, Kache (2015) conducted a Delphi study and recognised various opportunities and challenges of BDA adoption at the corporate and supply chain level. The participants of the Delphi study are experts in either one or both of Big Data Analytics and Supply chain management. SC professionals recognised visibility, transparency, and responsiveness as the main benefits of utilising BDA to manage Supply chain functions. The examples of BDA applications and their benefits discussed here are just the tip of the iceberg. It has enormous potential to transform traditional reactive supply chains into proactive data-driven supply chains. Sangari and Razmi (2015) have empirically verified that organisations find it difficult to establish cultural competence in supply chains. Various Key Performance Indicators (KPI) are used to measure the performance of supply chains. Organisations need to identify relevant KPIs in accordance with their business strategy. BDA can be used to monitor and measure KPI and assess the SC performance by comparing it with the benchmark (Jamehshooran et al. 2015b). Jamehshooran et al. (2015a) conducted an empirical study to determine the effect of supply chain analytics on supply chain performance. The authors discussed the use of Supply Chain Operations Research

(SCOR) model (which is an approved model of Supply Chain Council) as a systematic approach to determining supply chain performance. The SCOR model contains five key components: Plan, Source, Make, Deliver, and Return. Based on the Resource-based view and resource dependence theory, the authors have developed a conceptual framework and the key components are Plan analyses (PA), Source Analyses (SA), Make Analyses (MA), and Deliver Analyses (DA). On-Time Delivery, Quality, Cost, Reliability, and Flexibility are considered as dependent variables. Consequently, findings revealed a significant positive relationship between the use of supply chain analytics and supply chain performance. From the data analysis perspective, Jamehshooran et al. (2015a) have elaborated that the Plan process depends on analyses of data to predict market trends, customer requirements, and optimising resource utilisations. However, in comparison, the effect of analytics in the planning and manufacturing process is found to have a greater effect on SCP than SA and DA. Moreover, DA is found to have the least effect on SCP.

In a dynamic environment, strategic system design is vital to enhance supply chain performance and achieve a competitive advantage. Fawcett and Waller (2014) argued that Big Data and predictive analytics is one of the five-game chargers (others being additive manufacturing, autonomous vehicles, material science, and borderless supply chains) of supply chain design. Moreover, the authors consider four factors that may impede supply chain transformation into a better system that can co-create value. The four aspects that hinder supply chain transformation to co-create value are security concerns in supply chains, unsuccessful change management, lack of trust, and lack of understanding of CSR initiatives (Fawcett and Waller 2014). Hence, it can be argued that supply chain analytics could play a significant role in improving the performance of manufacturing firms.

2.6 Systematic literature review II - Absorptive capacity

Nowadays, absorptive capacity is considered as one of the renowned concepts in academic literature, which is developed as a consequence of perceiving Research and Development (R&D) as an important factor for firm performance. Cohen and Levinthal (1989, p.569) argued that a firm’s Research and Development (R&D) efforts, which contribute to the knowledge base, can produce two benefits: 1. Generating new information, 2. Increasing the ability of the firm to assimilate and exploit new information. This is in contradiction to the traditional belief of viewing new information generation as an only outcome of R&D efforts. Several authors such as Titon (1971), Evenson and Kislev (1975), Mowery (1983), and Allen (1984) are the foremost to observe this unique ability of firms to commercially exploit external knowledge (Lane et al. 2006). However, Cohen and Levinthal (1989, 1990) are the first to term this kind of ability as ‘absorptive capacity’, or in other words, the ‘learning capacity’ of an organisation, and are also the first to empirically verify it. Moreover, it is contended that the concept of Absorptive Capacity(ACAP) is adapted from the macroeconomics discipline (Kauppi et al. 2013). Ever since it is introduced, exploiting or commercialisation of external knowledge is perceived as a critical source of a firm’s competitive advantage and innovation. Cohen and Levinthal (1994) noted that firms that develop ACAP can exploit external information and flexibly adapt to the product’s market.

Due to intra-industry knowledge spillovers, firms could acquire new information, which diminished the incentive to invest in firms’ R&D efforts (Aghion and Jaravel, 2015). This traditional notion of R&D investment is changed after Cohen and Levinthal introduced the concept of ‘absorptive capacity’. Because “a firm's capacity to absorb externally generated knowledge depends on its R&D effort” (Cohen and Levinthal 1989). This signifies that the R&D efforts of an organisation can be an important antecedent to develop ACAP. Moreover, though ACAP is denoted as a capability, contrastingly it was proposed as “a function of the firm's level of prior related knowledge” (Cohen and Levinthal, 1990). The prior related knowledge, which is argued to increase the learning rate of an organisation, could include but is not limited to basic skills, shared language, knowledge about new technological developments, etc. It can be argued that this prior related knowledge is static, accumulated over a period, and enriched by R&D efforts. In the 1989 paper, significant emphasis is given to the prior knowledge base as it can enhance the learning of new concepts through associative learning mechanism. Besides the prior knowledge base, Cohen and Levinthal (1990) argued that an organisation’s prior learning experience can also influence the performance of the learning and knowledge transfer.

There are several factors that can have an effect on organisational absorptive capacity such as ‘Communication structure’ (centralised or organic) and ‘Knowledge structure’. Cohen and Levinthal (1990) argued that the flow of information via a centralised function won’t effectively link the external environment and suggested an organic communication structure for the effective flow of information. Further, all actors involved in the learning process should have overlapping knowledge, knowledge diversity, and “knowledge of who knows what, who can help with what problem, or who can exploit new information”(Cohen and Levinthal 1990). The critical knowledge or awareness of expertise that reside within and external to the organisation can complement diverse and overlapping knowledge structures. Besides, developing a buyer-supplier network relationship can help in leveraging individual ACAP which in turn could strengthen the organisation’s absorptive capacity (Cohen and Levinthal 1990). However, the authors argued that while a significant level of specialisation/concentration on overlapping knowledge can facilitate internal communication, it may hinder absorption of external knowledge from varied sources, due to its lack of diversity. So, ideally, it is important to develop effective internal and external relationships that can enhance an organisation’s knowledge structure, have partially overlapping knowledge and at the same time retain knowledge diversity. Referring to the initial discussions on the prior knowledge base, it is evident that prior knowledge should be somewhat ‘related’ to the new knowledge in order to assimilate it and, to a certain degree ‘diverse’ as well to facilitate effective and creative absorption of new knowledge.

Apart from the factors that may influence ACAP, two important features of absorptive capacity discussed are “the cumulativeness of absorptive capacity and its effect on expectation formation” (Cohen and Levinthal 1990). Accordingly, firms that have developed some ACAP can accumulate it more efficiently in their subsequent efforts, and can accurately anticipate or foresee commercial value of new technological knowledge. The ACAP feature of enabling accurate prediction of future technological advances are also advocated in their article “*Fortune Favors the Prepared Firm*” (Cohen and Levinthal 1994). Cohen and Levinthal (1990) argued that since ACAP is intangible in nature, it is difficult to quantify whether an organisation has attained its optimal level of ACAP or not.

Having discussed the background information on how ACAP is originally conceptualised, the reforms it has gone through warrants critical discussion as well. Over 25 years, the concept has gone through some significant reconceptualization. ACAP has been accepted and applied widely both in its original and reconceptualised form. In academic literature, there are various definitions and conceptualisation of

Absorptive Capacity (ACAP). Lane et al. (2006) critiqued Cohen and Levinthal (1990)'s work and argued that the paper deviated from the original conceptualisation of ACAP as 'capability' by modelling and testing it empirically using R&D intensity (a static resource) as a proxy, while R&D is in fact argued as an antecedent of ACAP.

From the introduction of the Relational view of the firm (Dyer and Singh 1998), it can be argued that researchers have started to perceive innovation as an inter-firm aspect, causing a paradigm shift from closed intra-firm innovation to open innovation. The prominent works of Dyer and Singh (1998), Lane and Lubatkin (1998) and Mowery et al. (1996), relevant to organisational learning, have shifted the focus of research from intra-firm to inter-firm knowledge transfer. Lane and Lubatkin (1998) are the first to critique and reconceptualise ACAP taking a relative or dyadic perspective to capture the role of both learning and teaching firms in the knowledge exchange process. This view is significantly different from Cohen and Levinthal (1990), who had articulated that firms with ACAP can equally learn from all external organisations. In addition, the original work of Cohen and Levinthal (1990, p.128) defined ACAP as "the ability of a firm to recognize the value of new, external information, assimilate it, and apply it to commercial ends is critical to its innovative capabilities". The immediate source of external knowledge for a firm is most probably its alliance partners. It is stated that "the ability of a firm to learn from another firm is jointly determined by the relative characteristics of the two firms, particularly the relationship between their knowledge-processing systems" (Lane and Lubatkin 1998). Thus, Lane and Lubatkin (1998)'s perspective sets importance on understanding the sources of knowledge, mechanisms that facilitate transfer and exploitation of it within the context of inter-firm alliances.

Further, Lane and Lubatkin (1998) claimed that the Relative ACAP consists of similarity in knowledge base, similarity in organisational structure and compensation practice, and dominant logic. These dimensions aimed to measure 'valuing new knowledge', 'assimilating new knowledge', and 'commercialising new knowledge'. The authors have used proxies to measure these dimensions. Relevance of knowledge base and joint research communities are used as proxies for the first and third dimensions. The firm's similarity with the alliance partner's organisational structure and compensation practices are used as a proxy to measure the ability to assimilate new knowledge from a partner firm. Further, the items used to measure

ACAP are referring to static resources established jointly by the organisations in alliance. Although not explicitly mentioned in (Lane and Lubatkin 1998), perhaps it can be argued that the concept of viewing ACAP from a process perspective (exploration, assimilation, and exploitation) originated from their work. They supplemented it empirically by finding that knowledge processing systems alias 'assimilation' play a major role in knowledge absorption.

Another important reconceptualization of ACAP is from Zahra and George (2002a), who claimed that ACAP is a dynamic capability. Unlike substantive or ordinary capability, the dynamic capability has the potential to induce organisational change. Zahra and George (2002a, p.186) define ACAP as "a set of organisational routines and processes by which firms acquire, assimilate, transform and exploit knowledge to produce a dynamic organisational capability". By recognising ACAP as a dynamic capability, it is argued that ACAP can lead to the development of other organisational competencies or substantive capability. Zahra and George (2002a) reasoned that the failure to capture multidimensionality and usage of varied definitions of ACAP by papers such as Mowery and Oxley (1995) and Kim (1995), had created the necessity to reconceptualise it. There are three important implications that can be derived from the definition of Zahra and George (2002): 1. ACAP is "embedded in a firm's routines and processes" (p.186), 2. The four dimensions (acquire, assimilate, transform and exploit) together can lead to the development of dynamic organisational capability, 3. The four dimensions are dependent and complement one another. Consequently, each subset of ACAP is defined as: 1. "Acquisition refers to the capability to identify and acquire externally generated knowledge that is critical to its operations" (p.189), 2. "Assimilation refers to the firm's routines and processes that allow it to analyse, process, interpret, and understand the information obtained from external sources" (p.189), 3. "Transformation denotes a firm's capability to develop and refine the routines that facilitate combining existing knowledge and the newly acquired and assimilated knowledge" (p.190), 4. Exploitation refers to firms ability "to refine, extend, and leverage existing competencies or to create new ones by incorporating acquired and transformed knowledge into its operations" (p.190).

Moreover, Zahra and George claimed that acquisition and assimilation are the fission of Potential ACAP (PACAP), and similarly, transformation and exploitation are of Realised ACAP (RACAP). The PACAP can facilitate the acquisition of valuable information but requires RACAP to exploit and translate it into value creation. However, PACAP alone cannot improve the firm's performance, the ratio of RACAP to PACAP has to be high in order to achieve it (Zahra and George 2002a).

Thus, previous studies such as (Lane and Lubatkin 1998) and (Dyer and Singh 1998) have provided implications for applying ACAP in the supply chain context. However, it can be argued that academics have started to apply it widely in a supply chain context following the contributions made by Zahra and George (2002), Malhotra et al. (2005) and Ettlie and Pavlou (2006), and its usage in a supply chain context has increased ever since. This is evident from the bibliometric analysis discussed in section 2.6.3.2. However, the contribution of ACAP theory to the supply chain literature is still ambiguous. Moreover, all the conceptualisations of ACAP discussed here have been critiqued in academic literature. For instance, Todorova and Durisin (2007) have criticised that Zahra and George's reconceptualization did not include the 'recognising the value' aspect of (Cohen and Levinthal 1990)'s original definition, and the misrepresentation of assimilation and transformation dimensions of ACAP. So, having discussed various conceptualizations and multifaceted nature of the ACAP construct, it is important to understand: 1. the antecedents and consequents of ACAP in a supply chain context, 2. in what ways ACAP is conceptualised to address supply chain issues. However, to the best our knowledge, there is no systematic literature review conducted relating to the context of ACAP and supply chain management. While Knoppen et al. (2015) reviewed journal papers on absorptive capacity in a supply chain context, their review takes the perspective of relationship learning and how processes and mechanisms facilitate learning and have a limited focus on the absorptive capacity construct.

2.6.1 Research methodology

The main purpose of this section is to elaborate on the systematic search of the literature, quality assessment, synthesis and validation of the findings. The method used to perform an SLR of the ACAP concept is similar to the method presented in section 2.2, and is well known and widely applied by several researchers such as Fahimnia et al. (2015) and Mishra et al. (2016). The six-step approach (defining the research question, determining the required characteristics of primary studies, retrieving a sample of potentially relevant literature, selecting the pertinent literature, synthesizing the literature, and reporting the results) discussed in (Durach et al. 2017) is adopted. Also, in line with (Tranfield et al. 2003), 'descriptive analysis' and thematic analysis was used to present the research findings. Consequently, the next step in the process is to retrieve a sample via a systematic search of a database.

2.6.2 Literature search

Scopus was used to source journal papers related to the topic, as it is the largest academic literature database that stores journal papers from more than 20,000 peer-reviewed journals (Fahimnia et al. 2015). A simple search using 'Absorptive capacity' as a keyword on Scopus, generated a sample of 3,918 research papers. Nevertheless, this research focuses only on the extracting a subsample of supply chain related papers from this larger sample. So, the keywords used to search journal papers were derived based on the research questions and contains a combination of 'absorptive capacity' and supply chain related terms such as supply management and operations management. In order to conduct a comprehensive search and include all relevant papers, 16 different combinations of keyword search were performed. Several inclusion and exclusion criteria were used to ensure the quality of the sourced papers. Only papers published in 'peer-reviewed journals' were included, and others from conference publications were excluded. In addition, only published papers and articles-in-press, that are in the 'English language' were included, while the rest were excluded. To avoid subjectivity and conduct an unbiased search, the inclusion and exclusion criteria were assessed autonomously by the researcher and the project supervisors (Tranfield et al. 2003). Consequently, Table 2.8 contains the combination of search terms used in this research, which resulted in a total of 142 papers, but after removing the duplicates there are 104 individual papers published from 1998 to 2017. Further, these 104 papers were read meticulously to identify those that are highly relevant to the topic. After reading the papers, studies that have used ACAP merely to describe or support their arguments without a central focus on it were excluded from further analysis. By employing these shortlisting criteria, the total number of articles was further reduced to 64. Among the 40 papers eliminated, 6 papers were not accessible and the others do not closely relate to the topic. In addition to the six-step approach mentioned earlier, the quality of the journal papers was evaluated by categorising based on the level of significance given to ACAP concept (

Table 2.9), using the approach of (Roberts et al. 2012). Further, the suitability of the final list of papers was verified and confirmed independently by the authors, ensuring intra- and inter-rater reliability, before further analysis.

Next, three different analyses were performed on the 64 journal papers. First, bibliometric citation analysis was conducted using 'BibExcel' software tool (Persson 2010), for the purpose of bibliometric and citation analysis. Second, content analysis was performed using frameworks developed by Roberts et al. (2012), which was previously used to review ACAP in information systems context. Accordingly, the journal papers were categorised based on 'conceptualisation', 'level of analysis', and the frequencies are presented in this research. Third, a thematic analysis was performed to identify research themes in the context of ACAP and SCM.

Table 2.8 Search Keywords used and initial search results

Search keywords (25sep2017)	Initial search results
"absorptive capacity" AND "supply chain"	52
"absorptive capacity" AND "operational performance"	8
"absorptive capacity" AND "operations management"	8
"absorptive capacity" AND "operations research"	2
"Operational absorptive capacity"	2
"absorptive capacity" AND "logistics"	12
"absorptive capacity" AND "supply management"	8
"absorptive capacity" AND "supplier innovativeness"	3
"absorptive capacity" AND "supplier innovation"	2
"absorptive capacity" AND "buyer-supplier"	12
"absorptive capacity" AND "supplier performance"	2
"absorptive capacity" AND "Supply network"	5
"absorptive capacity" AND "purchasing"	5
"absorptive capacity" AND "inter-organizational learning"	5
"absorptive capacity" AND "inter-organisational learning"	5
"absorptive capacity" AND "interorganizational learning"	6
"absorptive capacity" AND "interorganisational learning"	5
Total	142

Total (after removing duplicates)	104

Table 2.9 Journal papers categorised on a significance level of ACAP

	Referenced as a background or minor citation	**Provides theoretical support**	**Used in the hypothesis, proposition, or the research model**	**Forms the theoretical base of the article**	**Total**
Journal paper	(Knoppen et al. 2015)	(Meinlschmidt et al. 2016), (Mehmood et al. 2017), (Preston et al. 2017), (Bagchi et al. 2014), (Azadegan et al. 2008), (Jean et al. 2008), (Schiele 2007), (Narasimhan and Narayanan 2013), (Gammelgaard et al. 2011)	(John E Ettlie and Pavlou 2006), (Maike Scherrer-Rathje et al. 2014), (Cruz-González et al. 2015), (Saenz et al. 2014), (Pankaj C Patel et al. 2012), (Wei et al. 2015), (Chen et al. 2015), (Azadegan and Dooley 2010), (Liu et al. 2017), (Ambulkar et al. 2016), (Obayi et al. 2017), (Lawson and Potter 2012), (Azadegan 2011), (Kim and Lee 2015), (Revilla et al. 2013), (Chowdhury et al. 2017), (Setia and Patel 2013), (H Liu et al. 2013), (Kim et al. 2015), (Fayard et al. 2012), (Roldan Bravo et al. 2016), (Whitehead et al. 2016), (Lawson et al. 2015), (Lee and Song 2015), (Dobrzykowski et al. 2015), (Zhang et al. 2015), (Bellamy et al. 2014), (Lin 2014), (Abareshi and Molla 2013), (Schildt et al. 2012), (Arroyo-López et al. 2012), (Nagati and Rebolledo 2012), (Berghman et al. 2012), (Zacharia et al. 2011), (Jabar et al. 2011), (Arnold et al. 2010), (Tu et al. 2006), (Petroni and Panciroli 2002), (Riikkinen et al. 2016), (Cheng and Lu 2017), (Najafi Tavani et al. 2013), (Tavani et al. 2013), (Pihlajamaa et al. 2017), (Yang et al. 2017), (Liu 2012), (Malhotra et al.	(Khan and Nicholson 2014), (Liao and Marsillac 2015), (Tiep 2006), (Beheregarai Finger et al. 2014), (Mukherjee et al. 2000), (Hosseini and Khaled 2016)	

			2005), (Lane and Lubatkin 1998), (Kauppi et al. 2013)		
Frequency (%)	1 (1.56%)	9 (14.06%)	48 (75%)	6(9.37%)	64

2.6.3 Descriptive Analysis

2.6.3.1 Distribution of papers per year

As discussed earlier, the search strategy used had resulted in a total of 64 relevant journal papers. In the inclusion and exclusion criteria, the time limit is not set. However, the publication year of the final list of shortlisted journal papers ranges from 1998 to 2017. From Figure 2.19, it is apparent that there is a growing trend in terms of using ACAP in the supply chain context. Moreover, it can be argued that Malhotra et al. (2005) was the first to explicitly use ACAP in the context of supply chain, which leads to several publications related to the topic and also received a significant number of citations.

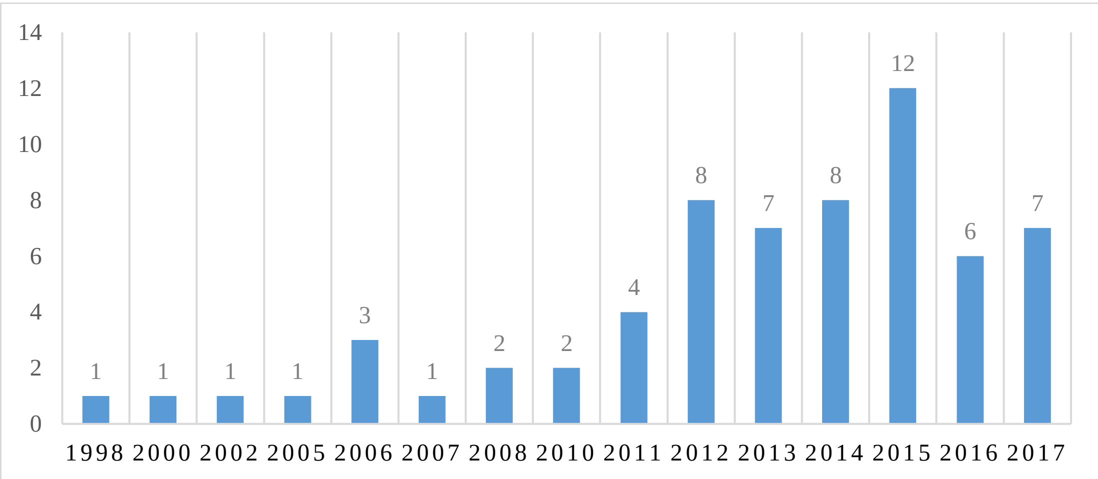


Figure 2.19 Distribution of research papers over the years

2.6.3.2 Top contributing authors

In terms of measuring the individual contributions of the researchers, the author’s name and the number publications in ACAP and SCM research domain were extracted using BibExcel tool. The evaluation was based on the number of publications, number of citations received by individual authors and their h-index score. Figure 2.20, reveals that authors Azadegan, Knoppen & Saenz have contributed three papers, respectively. But, when the number of citations received by the authors was considered (Figure 2.21), Lane, Lubatkin, Malhotra, Gosain and El Sawy had received the highest number of overall citations. However, h-index, calculated by combining citation count and a number of publication, shows that Azadegan, Saenz, and Knoppen are the top three contributors to this research domain. Hence, it can be recognised that Azadegan's contribution to this domain is considerably more significant than the others.

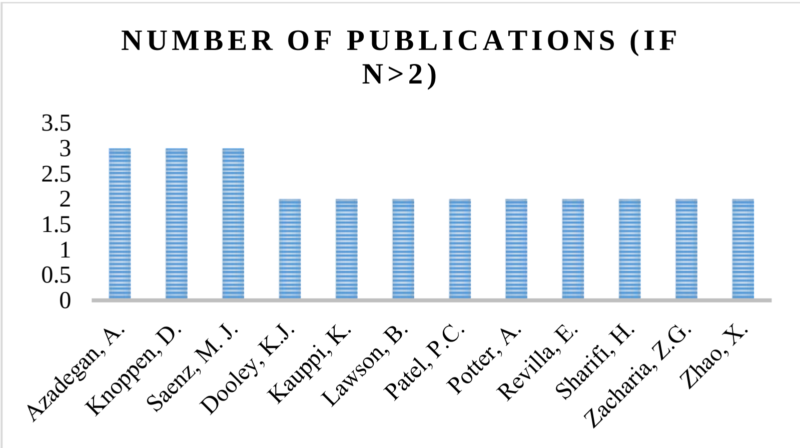


Figure 2.20 Top contributing authors based on the number of publications

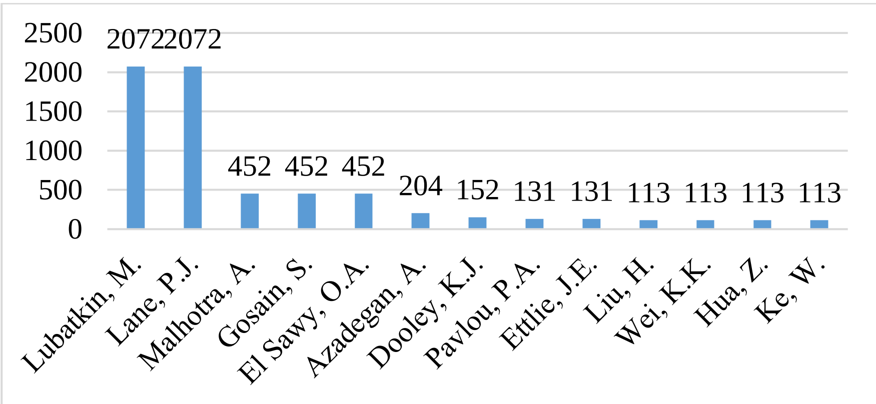


Figure 2.21 Top contributing authors based on the number of citations

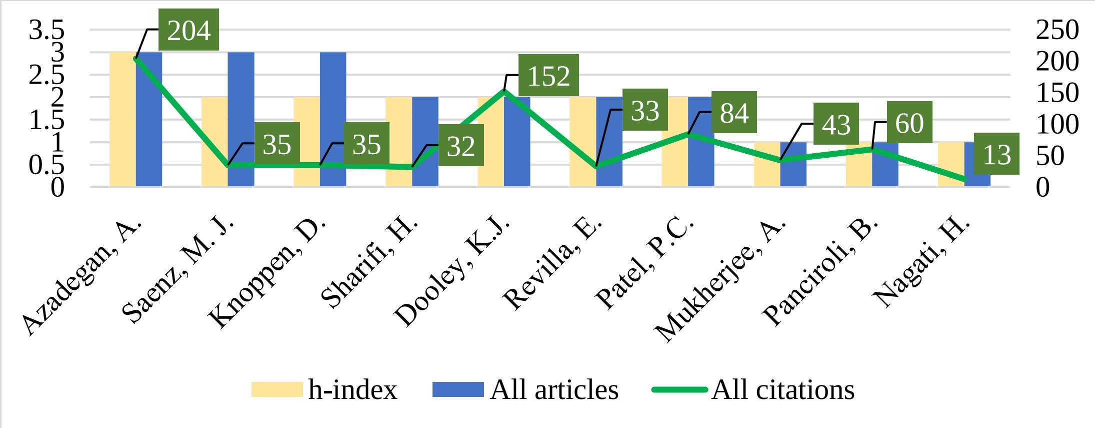


Figure 2.22 Top contributing authors based on h-index, number of publications and citations

2.6.3.3 Top cited papers

Among the 64 journal papers, (Lane and Lubatkin 1998), (Malhotra et al. 2005), and (Ettlie and Pavlou 2006) are the most highly cited papers (Table 2.10). It is well known that Lane and Lubatkin (1998) was the first to conceptualise relative absorptive capacity from a dyadic learning perspective. Malhotra et al. (2005), drawing from (S. Zahra and George 2002a) and (Dyer and Singh 1998), considered ACAP as a capability and measured their ability to exchange knowledge with a specific partner of the supply chain. Ettlie and Pavlou (2006) contended that ACAP is one of the significant components of inter-firm partners' dynamic capabilities.

Table 2.10 Top cited journal papers

Journal papers	Citation count	Journal papers	Citation count
(Lane and Lubatkin 1998)	2072	(Azadegan 2011)	52
(Malhotra et al. 2005)	452	(Schiele 2007)	50
(Ettlie and Pavlou 2006)	131	(Jean et al. 2008)	45
(H Liu et al. 2013)	113	(Mukherjee et al. 2000)	43
(Tu et al. 2006)	111	(Schildt et al. 2012)	39
(Azadegan and Dooley 2010)	87	(Bellamy et al. 2014)	33
(Zacharia et al. 2011)	85	(Narasimhan and Narayanan 2013)	30
(Azadegan et al. 2008)	65	(Liu 2012)	27

(Pankaj C Patel et al. 2012)	60	(Saenz et al. 2014)	26
(Petroni and Panciroli 2002)	59	(Fayard et al. 2012)	26

2.6.3.4 Affiliation statistics

The affiliations of all the first authors were extracted using BibExcel. The statistics on contributions made by countries around the world were analysed. From Figure 2.23 and Figure 2.24, it is evident that the majority of publications related to ACAP and SCM came from United States of America (USA), United Kingdom and Spain. Contributions from South America, Asia, Middle East and Australia are the least. Moreover, the researchers’ attention to the regional context was also explored. Yet again, the context of the majority of studies focuses on firms in the USA, China, and Europe (Figure 2.25).

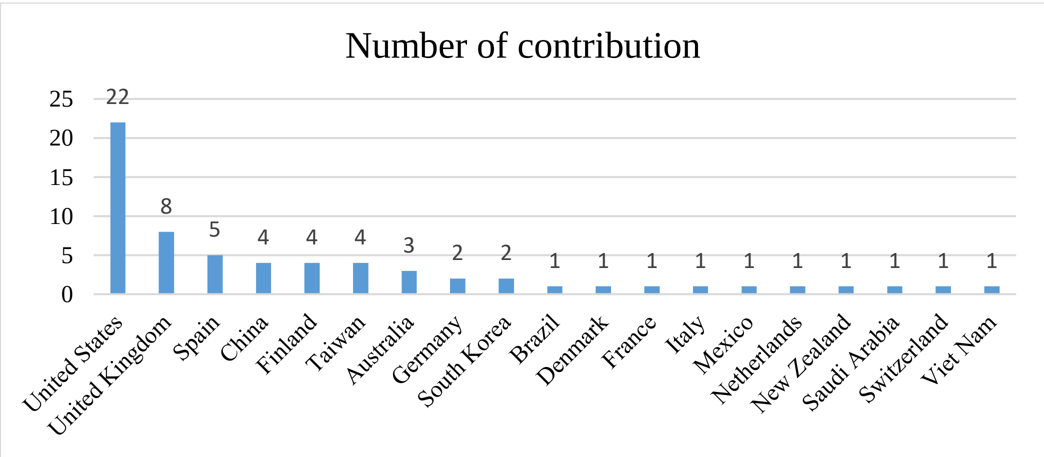


Figure 2.23 Classification of journal papers based on affiliation

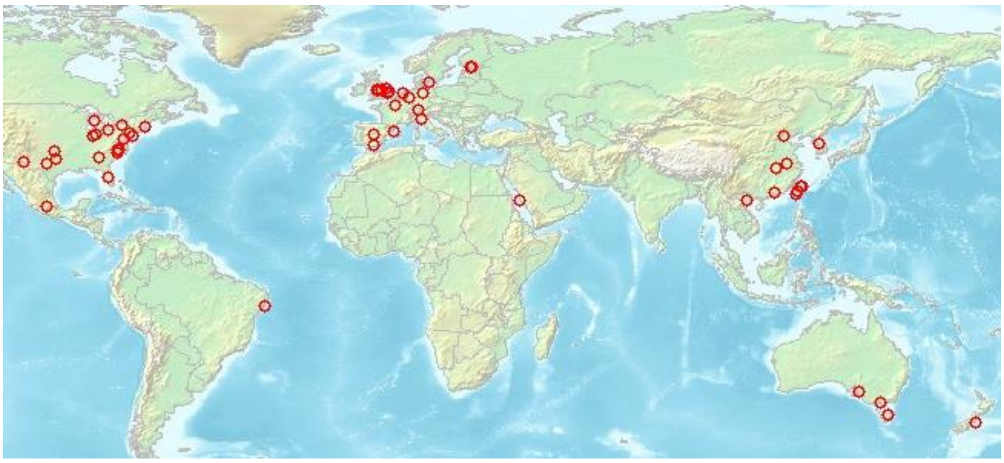


Figure 2.24 Visualisation of sources of publications

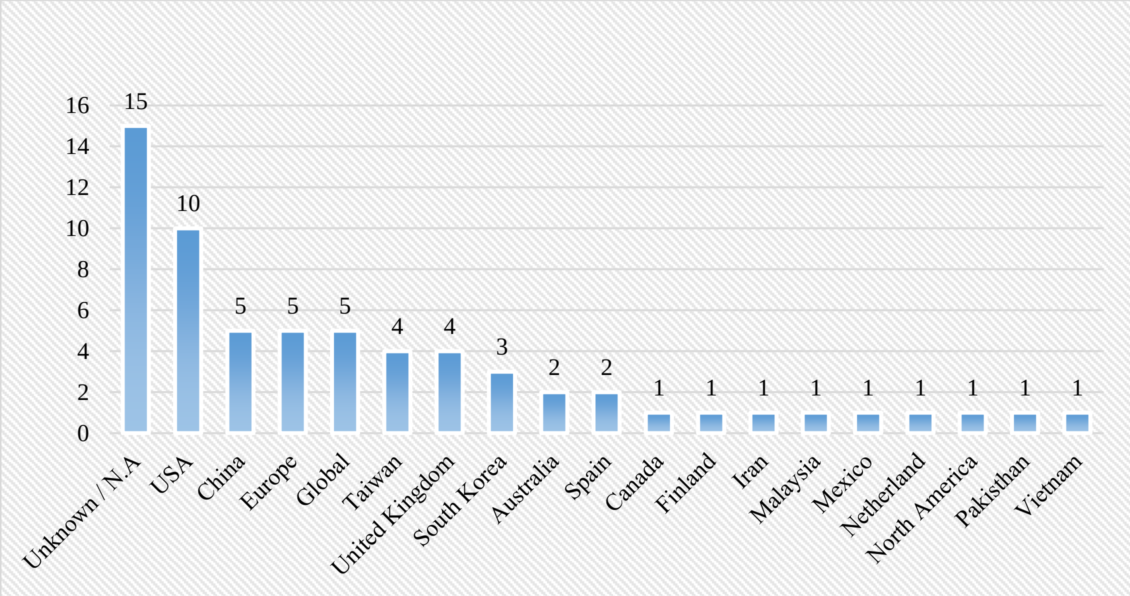


Figure 2.25 Classification of journal papers based on regional context of the study

2.6.3.5 Keyword statistics

It can be argued that analysing keywords mentioned in the journal papers can give useful insights on topics closely related to ACAP and the SCM domain. When the keyword statistics was performed, it was identified that the terms 'absorptive capacity' and 'supply chain management' are the most frequently used. Besides, researchers had also used supply chain related terms such as 'supply chain collaboration', 'buyer-supplier relationship' and 'supplier involvement', indicating strategic collaboration and relationship building as an important research domain associated with ACAP. Further, terms such as 'innovation', 'product development', 'knowledge transfer' and 'performance' can be viewed as the most common ACAP outcomes.

Table 2.11 Keyword statistics - ACAP and SCM

Keyword	Frequency	Keyword	Frequency
Absorptive capacity	45	Organizational learning	5
Supply chains/Supply chain management	17	Knowledge acquisition	4
Knowledge management	9	Societies and institutions	4
Innovation	9	Knowledge transfer	4
Industry	8	Supplier involvement	4
Collaboration /supply chain collaboration	8	Operations strategy	4
New product development	6	Structural equation modelling	4
Buyer-supplier relationships	6	Performance	3
Knowledge-based systems	5	Technology adoption	3

2.6.3.6 Types of research methods

As it is important to explore the research methodology used by supply chain researchers to investigate the phenomenon of ACAP and SCM, the content of the journal articles was quantitatively reviewed to find the most frequently used research methods. Findings suggest that more the half of the studies have used purely survey-based approach (Figure 2.26). It can be argued that the use of survey method is an indication of the well-developed nature of the ACAP construct because most of these studies have used the construct in hypothesis development. Moreover, a mixed methods approach is the second most widely used method. However, this research domain particularly lacks in producing qualitative research such as interviews and case studies.

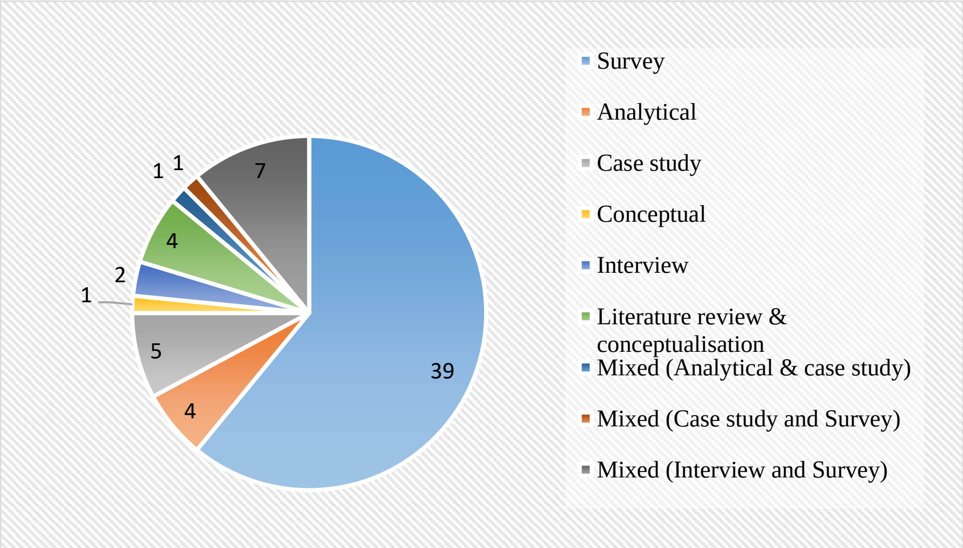


Figure 2.26 Classification of selected papers based on research methods

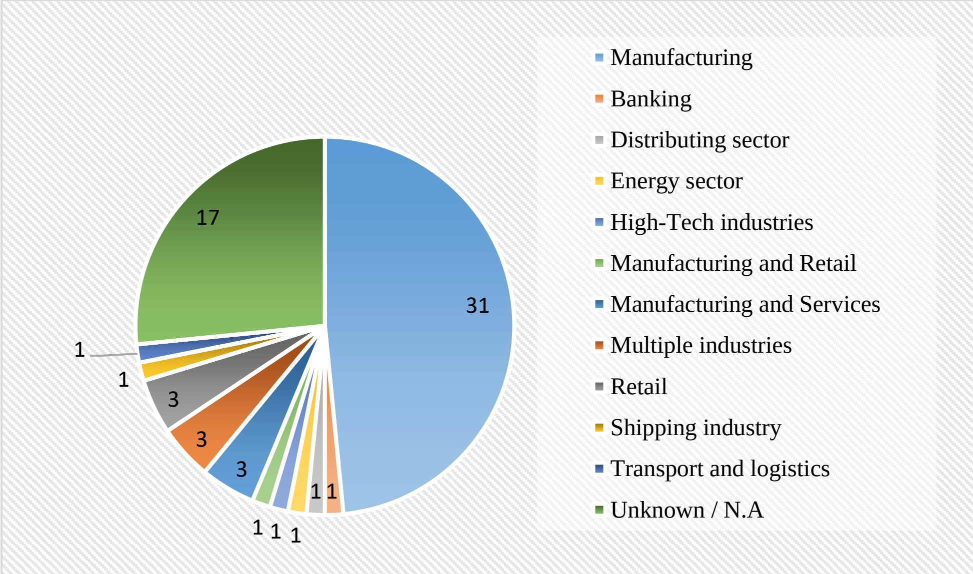


Figure 2.27 classification of journal papers based on industry focus

The final descriptive analysis involves finding the top contributing journals to the topic of discussion. The findings can lead to recognising the emphasis given to this domain by leading journals. Figure 2.28 highlights that the International Journal of Operations and Production Management (IJOPM) has published around 10 papers investigating the role of ACAP in supply chain context, followed by the Journal of Operations Management (JOM) and the Journal of Supply Chain Management (JSCM).

Figure 2.28 Contributions from different journals

2.6.4 Content Analysis

2.6.4.1 Conceptualisation of ACAP in SCM

Lane et al. (2006) critically reviewed the ACAP construct and recommended that the construct is 'reified' and lacks cohesion in conceptualising it. In this section, the journal papers were examined to identify the conceptual understanding of supply chain researchers. Among supply chain researchers, ACAP was conceptualised as both asset and capability, suggesting there is no cohesion in conceptualisation. An asset can be tangible or intangible in nature that an organisation has access to and control over it. In contrast, 'organisational (or substantive) capability' is referred as a set of routines and processes that the organisations use, and it is further distinguished from 'dynamic capability' which denotes the ability of firms to reconfigure substantive capabilities (Roberts et al. 2012).

More than 50 % of supply chain researchers have conceptualised ACAP as a capability, but still, a considerable number of researchers have recognised it as an asset. For instance, Wei et al. (2015) conceptualised ACAP as an asset and argued that the firm's prior knowledge of technology can enhance their ability to assimilate supply chain technology such as RFID. Kim et al. (2015) considered 'overlapping knowledge' and 'cognitive congruence' with supply chain partner as ACAP. It can be argued that supply chain partners having a similar knowledge base is a representation of an asset, as it is static in nature. Bellamy et al. (2014) conceptualised it as an asset and used 'R&D intensity' as a proxy to measure ACAP. Similarly, Schildt et al. (2012) used 'R&D intensity' and 'technology similarity' measured through similarity of patents between supply chain partners as a proxy for ACAP. However, Roberts et al. (2012) argued that ACAP must be conceptualised as a capability and raised the concern of construct validity if ACAP is perceived as an asset.

On the other hand, several studies such as (Cheng and Lu 2017) (Yang et al. 2017) (Riikkinen et al. 2016) and (Malhotra et al. 2005) have considered ACAP as a capability. ACAP conceptualised as capability is found to have a positive influence on proactive and reactive dimensions of supply chain resilience (Cheng and Lu 2017). Yang et al. (2017) have indicated 'potential absorptive capacity', drawing from Zahra and George (2002a), as a capability and found it can strengthen the relationship between operational improvement practices and operational efficiency. Riikkinen et al. (2016) also adopted a similar approach and explained the relationship between the potential and the realised ACAP of purchasing department sustainability and firm performance. Moreover, only a few studies have not explicitly conceptualised ACAP either as an asset or capability. These findings show that there is a varied understanding of the concept among supply chain researchers.

Table 2.12 Absorptive capacity and its conceptualizations

Conceptualised as	**Frequency**	**Journal papers**
Asset (%)	18 (28.12%)	(Kauppi et al. 2013) (Wei et al. 2015) (Lane and Lubatkin 1998) (Kim et al. 2015) (Fayard et al. 2012) (Bellamy et al. 2014) (Lin 2014) (Bagchi et al. 2014) (Schildt et al. 2012) (Azadegan et al. 2008) (Tu et al. 2006) (Petroni and Panciroli 2002) (Mukherjee et al. 2000) (Tavani et al. 2013) (Pihlajamaa et al. 2017) (Jabar et al. 2011) (Liu 2012) (Hosseini and Khaled 2016)
Capability (%)	37 (57.8%)	(Ettlie and Pavlou 2006) (Scherrer-Rathje et al. 2014) (Cruz-González et al. 2015) (Saenz et al. 2014) (Patel et al. 2012) (Chen et al. 2015) (Liu et al. 2017) (Ambulkar et al. 2016) (Obayi et al. 2017) (Meinlschmidt et al. 2016) (Lawson and Potter 2012) (Azadegan 2011) (Liao and Marsillac 2015) (Kim and Lee 2015) (Revilla et al. 2013) (Chowdhury et al. 2017) (Mehmood et al. 2017) (Setia and Patel 2013) (Liu et al. 2013) (Roldan Bravo et al. 2016) (Whitehead et al. 2016) (Lawson et al. 2015) (Dobrzykowski et al. 2015) (Zhang et al. 2015) (Beheregarai Finger et al. 2014) (Abareshi and Molla 2013) (Arroyo-López et al. 2012) (Nagati and Rebolledo 2012) (Berghman et al. 2012) (Zacharia et al. 2011) (Arnold et al. 2010) (Jean et al. 2008) (Riikkinen et al. 2016) (Cheng and Lu 2017) (Yang et al. 2017) (Malhotra et al. 2005)
Not explicitly conceptualized (%)	9 (14.06%)	(Azadegan and Dooley 2010) (Khan and Nicholson 2014) (Knoppen et al. 2015) (Tiep 2006) (Preston et al. 2017) (Lee and Song 2015) (Schiele 2007) (Narasimhan and Narayanan 2013) (Gammelgaard et al. 2011)
Total	64	n/a

2.6.4.2 Level of analysis

ACAP is argued to be a multilevel construct (Roberts et al. 2012). Cohen and Levinthal (1990) have discussed the similarities and variances between individual and organisational ACAP. They suggested that ACAP of an organisation depends on its individual employees' ACAP, but it is not simply an aggregation of individual ACAP, rather it links across individual ACAP. At the individual level, ACAP can be developed through personnel training. However, organisations should focus on the role of individuals or gatekeepers at the interface with the external environment to seek new information and, at the same time, manage communication networks within subunits of the organisation. Developing ACAP at the organisational level can be complex as it involves contemplating intricate organisational mechanisms and factors that can facilitate it.

Subsequently, the shortlisted journal papers were categorised into five levels based on how ACAP is applied in SCM research into: 1) individual, 2) group/team, 3) organisational, 4) inter-organisational, and 5) country level. From Table 2.13, it is evident that only 3% of papers measure ACAP at the individual level, around 11 % measure it at group/team level, 40.6 % measure it at the organisation level, and just 1.5% of papers measure ACAP at the country level in the context of SCM. Moreover, the highest of all, i.e. 43.75% of papers measure ACAP at the inter-organisational level.

Two papers by Ambulkar et al. (2016) and Kauppi et al. (2013) have measured ACAP at the individual level. Kauppi et al. (2013) argued that the buyer's and manager's prior relevant knowledge (i.e. Individual asset) enhances their ability to use e-procurement tools effectively to contribute to firm performance. Whereas, Ambulkar et al. (2016) conceptualised it as an individual's capability and stated that "the individual's ability to acquire external knowledge, disseminate it and exploit it to fulfil firm's business objectives (p.1400)" can be referred to as individual ACAP. Further, it is argued that supply chain managers with a high level of ACAP can proactively acquire information needed to mitigate supply chain risks under environmental uncertainty.

Moreover, few researchers have used ACAP at the group/team level. Setia and Patel (2013) and Patel et al. (2012) have measured it specific to the operations department. The ability of the operations department to acquire operational knowledge from external sources, assimilate and exploit it is called operational absorptive capacity (Setia and Patel 2013). Patel et al. (2012) defined it as "the ability of a firm's operational units to acquire, assimilate, transform, and exploit knowledge from the operations environment". Moreover, Setia and Patel (2013) conceptualised it as a capability containing potential and realised operational ACAP, following the approach of (Zahra and George 2002a), and information technology is found to be an antecedent of operational absorptive capacity. It is essential to develop operational ACAP to cope effectively with the uncertain environment (Pankaj C Patel et al. 2012; Setia and Patel 2013). Similarly, a service team's ACAP is found to positively influence operational performance (Yang et al. 2017), and the realised ACAP of a purchasing department is found to have a positive effect on social sustainability practices and economic performance (Riikkinen et al. 2016). Schiele (2007) found out that a high maturity of the firm's purchasing function will increase absorptive capacity and be able to provide cost benefits. Mukherjee et al. (2000), perceived ACAP as an asset, i.e., prior manufacturing routines, and argued that production lines use this prior routine in a static manner, unlike dynamic routines which require flexible adaptation of routines at the varying time frame. They further found out that the performance of the production line will decline when the newly introduced routine or knowledge is beyond the firm's capacity to process it. In other words, if an organisation tries to acquire and assimilate new knowledge that is not related to the prior knowledge base, it will result in a decline in performance.

Similarly, if the authors have not measured ACAP at individual or groups/team level or in relation to external partners, then the journal papers were classified into the organisational level. Moreover, if either the author's focus or the study participants' response is related to a particular organisation, then it was classified under 'organisational level' category. Consequently, 40.6 % (26 papers) measured ACAP at the organisational level. A high proportion of these papers conceptualised ACAP as a dynamic capability, as it is inherently considered as an organisational-level construct (Roberts et al. 2012). For instance, Hosseini and Khaled (2016) conceptualised ACAP, in the context of supply chain disruption, as the ability of supplier organisations' to absorb the shocks and withstand disruptive events. Azadegan and Dooley (2010) focused on a manufacturer's ACAP and found out that it can influence the effect of a supplier's innovativeness on a manufacturer's performance. On the other hand, Arroyo-López et al. (2012) measured the role of a supplier's ACAP on the success of supplier development programs. They conferred that it is essential to evaluate a supplier's ACAP before engaging in a relationship to ensure that suppliers have the capability to absorb new knowledge and for development programs to be successful. Liu et al. (2013) studied manufacturing organisations' ACAP driven by information technology to find that it has a significant influence on developing supply chain agile capabilities.

Lastly, there are 28 (43.75%) papers classified into the inter-organisational level. It is evident that supply chain researchers have used ACAP mostly at inter-organisational level. This is in contrast to the findings of (Roberts et al. 2012), which is in the context of Information Systems research. The majority of research under this category has adopted Relative ACAP (Lane and Lubatkin 1998), which conceptualised knowledge and structural similarity between the learning and the teaching firm as Relative ACAP. Kim et al. (2015) used the relative absorptive capacity conceptualisation of Lane and Lubatkin (1998). Further, the similarity in knowledge and similarity in cognitive/social context between buyer-supplier is considered as dimensions of absorptive capacity.

In Saenz et al. (2014), the suppliers' ACAP from a process perspective is measured to understand in what ways suppliers and buyers jointly make efforts to share information and exploit it. Similarly, in (Liao and Marsillac 2015), the structure of the External Knowledge Acquisition construct, developed using ACAP theory, exemplifies the joint effort between the buyer and the supplier to absorb external knowledge. Revilla et al. (2013) argued that ACAP is not a firm level but an inter-organisational phenomenon.

Unlike Saenz et al. (2014), Lawson and Potter (2012) measured the ACAP of buyers in terms of absorbing knowledge from their supplier, from a buyer's perspective. On the other hand, Revilla et al. (2013), conceptualised it as the ability to absorb knowledge from a specific buyer from a supplier's perspective. Similarly, Lawson et al. (2015) conceptualised relationship-specific ACAP focusing on buyers absorbing knowledge from strategic suppliers with specific technical knowledge. Although Lawson and Potter (2012) and others discussed here did not measure it from a joint effort perspective, it represents a dyadic inter-firm relationship hence is categorised into the inter-organisational level. In Whitehead et al. (2016), the knowledge distribution capability of the source firm is measured from a recipient perspective.

Moreover, Roldan Bravo et al. (2016) measured the ACAP of supply networks from the buyer's perspective, as a representative of ACAP among the members of the supply network. Similarly, Dobrzykowski et al. (2015) measured focal firms' and supply chain members' ability to acquire information using electronic tools, assimilate, transform and apply it to improve operational performance. Besides, a very small proportion of researchers have applied ACAP at the country level in the context of the supply chain.

Table 2.13 Absorptive capacity and its level of analysis

Levels*	Frequ- -ency	Journal papers
Individual (%)	2	Buyer and Manager competence (Kauppi et al. 2013), supply chain managers' ACAP (Ambulkar et al. 2016)
Group/team (%)	7	Operations unit (Patel et al. 2012), transport system (Mehmood et al. 2017), Operational department (Setia and Patel 2013), purchasing department (Schiele 2007), production line (Mukherjee et al. 2000), purchasing function (Riikkinen et al. 2016), Service team (Yang et al. 2017)
Organizational (%)	26	(Ettlie and Pavlou 2006), (Hosseini and Khaled 2016), (Maike Scherrer-Rathje et al. 2014), (Cruz-González et al. 2015), (Wei et al. 2015), (Chen et al. 2015), (Azadegan 2011), (Kim and Lee 2015), (Chowdhury et al. 2017), (Liu et al. 2013), (Lee and Song 2015), (Beheregarai Finger et al. 2014), (Bellamy et al. 2014), (Lin 2014), (Abareshi and Molla 2013), Supplier organisations ACAP (Arroyo-López et al. 2012), (Berghman et al. 2012), (Zacharia et al. 2011), (Azadegan et al. 2008), (Jean et al. 2008), (Tu et al. 2006), (Petroni and Panciroli 2002), (Tavani et al. 2013), (Gammelgaard et al. 2011), Manufacturer's Absorptive capacity (Azadegan and Dooley 2010)
Inter-organizational (%)	28	Buyer-Supplier ACAP (Saenz et al. 2014), Buyer-supplier ACAP (Lane and Lubatkin 1998), Buyer-Supplier relationship ACAP (Liu et al. 2017), Buyer-Supplier ACAP (Obayi et al. 2017), Buyer-supplier Absorptive capacity & Desorptive capacity (Meinlschmidt et al. 2016), Buyer-Supplier ACAP in Knowledge transfer (Khan and Nicholson 2014) , Supplier – buyer knowledge transfer (Lawson and Potter 2012), Buyer-supplier relationship ACAP (Liao and Marsillac 2015), Buyer-supplier ACAP (Knoppen et al. 2015), Buyer-supplier ACAP from supplier's perspective (Revilla et al. 2013), Buyer-supplier knowledge transfer (Tiep 2006), Buyer-supply relative ACAP (Similarity) (Kim

		et al. 2015), Focal firm and partner specific ACAP (Fayard et al. 2012), Buyer's and supply network's absorptive and desorptive capacity (Bravo et al. 2016), ACAP of recipient and distributive capability of source (Whitehead et al. 2016), Buyer-supplier knowledge transfer from supplier perspective (Preston et al. 2017), Supplier-buyer knowledge transfer from buyer's perspective (Lawson et al. 2015), ACAP at supply chain (Dobrzykowski et al. 2015), ACAP is both relation-specific and firm-level (Zhang et al. 2015), Relative absorptive capacity (Schildt et al. 2012), relative absorptive capacity (Nagati and Rebolledo 2012), Relative ACAP-similarity with partners (Jabar et al. 2011), ACAP in relation with supply chain partner (Arnold et al. 2010), ACAP in supply chain relationship –Joint effort (Cheng and Lu 2017), Relationship specific ACAP (Malhotra et al. 2005), Supplier-buyer (low R&D intensity) knowledge transfer (Pihlajamaa et al. 2017), ACAP at Supply network (Narasimhan and Narayanan 2013), Relation specific ACAP (Liu 2012)
Country level	1	(Bagchi et al. 2014)
Total	64	n/a

* 1. Organizational level- Studies solely focus on ACAP of an organisation as a unit of analysis 2. Inter-organizational level- a. Studies concentrating on AC of partners measured from focal firm's perspective or vice-versa, b. ACAP of a supply chain.

2.6.5 Thematic Analysis

Thematic analysis is a systematic process of categorising the content of the text and identifying relationships among the categories Berg(1995) as cited in (Lane et al. 2006). The content of the selected journal papers was reviewed to identify important themes emerging from it. Consequently, four themes were identified and listed in the Table 2.14 below.

Table 2.14 Research themes and key references

Themes	Key references
Technology-driven ACAP	(Jean et al. 2008) (Liu et al. 2013)(Kauppi et al. 2013)(Wei et al. 2015)
Relation-specific ACAP	(Saenz et al. 2014) (Lane and Lubatkin 1998) (Liu et al. 2017) (Obayi et al. 2017) (Revilla et al. 2013) (Tiep 2006) (Kim et al. 2015)

Knowledge-specific ACAP	(George et al. 2001) (Abareshi and Molla 2013) (Lee and Song 2015) (Ambulkar et al. 2016), (Setia and Patel 2013)
Ambidextrous ACAP	(Whitehead et al. 2016) (Meinlschmidt et al. 2016)

2.6.5.1 Technology-Driven Absorptive capacity

Malhotra et al. (2005) reiterated the view of Zahra and George (2002b) and suggested that ACAP is an IT-driven capability. Inter-organisational information systems deployed at the interface of the supply chain can enhance the ability of firms to acquire knowledge from the external environment (Malhotra et al. 2005). While the role of technology, in general, can provide inferences to break down various aspects of developing ACAP, it is still unknown to what extent ACAP is driven by technology in the context of supply chain research. In particular, the consensus is yet to arrive whether ACAP should be considered as an antecedent or a consequent of supply chain related technology. Thematically, there are two categories of researchers identified from the systematic review, one set of researchers considers ACAP as an antecedent and the others consider it as a consequent in relation to technology. In general, when it comes to assimilating new technology, ACAP is conceptualised as an asset (prior knowledge base), an antecedent of it. However, integration and practice of technology or exploitation of technical capabilities are perceived as an antecedent of developing dynamic ACAP.

For instance, Wei et al. (2015) argued that static ACAP is the antecedent of technology assimilation. Whereas, External Electronic Integration and IT capabilities are regarded as an antecedent of ACAP (Fayard et al. 2012; Jean et al. 2008; John E Ettlie and Pavlou 2006). The difference lies in the way these papers conceptualise ACAP. As discussed earlier, it is perceived both as an asset in (Wei et al. 2015) and capability in (Fayard et al. 2012) in the context of technology. Moreover, ACAP is considered as an antecedent not only for assimilating new technology but also for assimilating innovative manufacturing practices accomplished through utilisation of technology(Tu et al. 2006). Kauppi et al. (2013) found out the prior knowledge of buyers and managers is important to realise the influence of e-purchasing tools such as e-sourcing, e-process, and e-transaction on process cost and purchasing price.

ACAP has to be combined with e-purchasing tools to improve performance. Liu et al. (2013) argued that IT capabilities can help firms' to improve ACAP, and it is found out the IT-driven ACAP can enhance operational capabilities such as supply chain agility and firm performance. Moreover, Liu et al. (2013) viewed ACAP as dynamic capabilities, which refers to "a firm's ability to integrate, build, and reconfigure internal and external competencies"(Teece et al. 1997). Moreover, Liu et al. (2013) viewed dynamic capabilities from the hierarchical perspective and suggested that ACAP is a higher-order capability enhanced by lower-order IT capabilities. Similarly, Setia and Patel (2013), adopted Zahra and George (2002a)'s conceptualisation and suggested that information systems capability can help organisations to develop operational ACAP to acquired operations related knowledge. Whereas, Lin (2014) suggested that e-SCM, unlike intra-organisational systems, has to be jointly adopted by organisations in the supply chain. Further, it is argued that a firm's ability to absorb knowledge from a supply chain environment on how to adopt e-SCM can significantly promote adoption of e-SCM technologies deeper into the organisation.

2.6.5.2 Relationship-specific absorptive capacity

The key focus of researches under this theme is the knowledge transfer between partners in the supply chain relationship. It is noticed that there are two streams of research under this category in which ACAP is used to explain the underlying mechanisms of: 1. Buyer transferring knowledge to its suppliers, and 2. Buyer absorbing knowledge from its suppliers.

Kim et al. (2015) found that the supplier's relative absorptive capacity and trust on buyers significantly moderates the relationship between Buyer-driven knowledge transfer and supplier's operational performance. However, supplier innovativeness negatively moderates the relationship, which implies only suppliers that have less innovativeness and high trust are highly motivated to learn from buyers. Tiep (2006) categorised ACAP into specific and general absorptive capacity. Specific absorptive capacity is developed by local suppliers after establishing a relationship with the assembler. Establishment of buyer-supplier relationship helps suppliers to develop their buyer specific absorptive capacity. Further, according to Khan and Nicholson (2014), the increase in the level of a supplier's absorptive capacity from low to high is in correlation with different stages of the supplier development process such as qualification stage, evaluation stage, and an interactive stage. However, perception gap between buyers and suppliers is found to be a significant barrier for improving supplier's absorptive capacity (Khan and Nicholson 2014). Relative ACAP is considered as an antecedent of knowledge transfer from customer to suppliers (Nagati and Rebolledo 2012). Nagati and Rebolledo (2012) measured dimensions of relative ACAP as 'knowledge overlapping routines' and 'overlapping knowledge bases', and found out that the presence of similar knowledge base (i.e. asset) does not influence knowledge transfer. However, establishing knowledge sharing routines between the customer and suppliers will certainly enhance knowledge transfer and improve suppliers' performance. Jabar et al. (2011) argue that ACAP is one of the factors in enabling organisational learning and allowing technology transfer from a supply chain partner. Other factors such as the nature and type of the strategic technology alliance and the learning environment are complementary to it.

Unlike other studies which deal with the buyer to supplier knowledge transfer, there are studies under this theme which address the phenomenon of knowledge transfer occurring from the supplier side to buyers. Lawson et al. (2015)'s study focuses on strategic suppliers with specific knowledge. The authors empirically validated that a buyer's ability to absorb knowledge from strategic suppliers (relationship-specific ACAP) can enhance the performance of new product development. Lawson et al. (2015), in addition, argued that suppliers may lack the capability to transfer knowledge across inter-firm boundaries, and in such cases, the development of a relationship-specific ACAP could help them to transfer knowledge to buyers. Relationship-specific ACAP can act as a substitute to suppliers' lack of technical capabilities and provide mechanisms to transfer knowledge from suppliers to buyers. However, Lawson et al. (2015) found out that when suppliers technical capabilities are high, there is no significant link between relationship-specific ACAP and new product development performance, and the importance given to the development relationship-specific ACAP has to be moderated accordingly. Pihlajamaa et al. (2017) have found evidence from a case study that through effective supplier management capabilities, a buying firm can get benefits from supplier innovation, despite the low R&D intensity of buying firm. This indicates that in a supply chain relationship, firms with low absorptive capacity can get innovation benefit via establishing supplier specific capabilities to transfer knowledge from suppliers to buyers.

Schildt et al. (2012) addressed the temporal dynamics involved in developing ACAP specific to supply chain relationships. The authors argued, in the context of buyer-supplier alliances, that having a high level of ACAP increases the learning rate only at the later stage of the alliance, but it weakens the learning rate during the initial stages. Arnold et al. (2010) conceptualised ACAP into two dimensions: 1. Operational efficiency of their supply chain trading partner, 2. Knowledge creation capability- Ability of the supply chain trading relationship to enable knowledge creation within the organisation. It is found out that the buyers' commitment to its suppliers is high if suppliers have a high level of absorptive capacity. Moreover, a joint effort of managing supply chain knowledge by the focal firm and its partners can improve supply chain competencies such as resilience (Cheng and Lu 2017).

2.6.5.3 Knowledge-specific absorptive capacity

Researchers under this theme focus on acquiring a specific type of knowledge from the external environment. Unlike general ACAP and relation-specific ACAP, the emphasis is given to supply chain knowledge and its characteristics. It can be argued that knowledge is the key element of focus in the entire debate on the processes, capabilities, and mechanisms involved in managing and creating value from it. Polanyi (1967) is the first to articulate that knowledge can be characterised into tacit and explicit (Schoenherr et al. 2014). Explicit knowledge is codified and easy to communicate, but tacit knowledge is engrained, subjective and mostly reside within individuals and often difficult to communicate (Sivakumar and Roy 2004). This kind of tacit knowledge is considered by organisations as a valuable asset for competitive advantage. A supply chain is inundated with both these knowledge forms, and organisations make efforts to develop capabilities to acquire it. Moreover, Murovec and Prodan (2009), based on several pieces of evidence, suggested that ACAP is dependent on the knowledge characteristics and organisations require high absorptive capacities in terms of scientific knowledge compared to knowledge from other sources such as customers.

Researchers have discussed about the role of ACAP in acquiring specific-knowledge such as operations knowledge and environment knowledge. For instance, Setia and Patel (2013) place emphasis on developing ACAP to acquire operations knowledge. The authors argued that ACAP would enable operations department to acquire knowledge from their operational environment and would allow them to monitor changes in customer demands. Transforming and applying such knowledge can support organisations to adapt processes and reconfigure or create new products and services. Information systems capability is found to be an antecedent of developing operational knowledge specific to ACAP, but it is recognised that IS strategy should be aligned with the business strategy in order to realise the potential of operational knowledge. In Fayard et al. (2012), the ability of both the focal firm and the partner to transfer and receive knowledge about cost information is measured. ACAP is conceptualised into three dimensions: knowledge seeking, communication network, and communication climate. Electronic integration is an important source for enabling ACAP, which can improve inter-organisational cost management. Similarly, Chen et al. (2015) and Abareshi and Molla (2013) have conceptualised ACAP as the ability to absorb environmental information from external sources. (Abareshi and Molla 2013) found out that the concept of Green ACAP can significantly improve green performance. Chen et al. (2015)'s study also supports this finding and provided further empirical evidence on its positive influence on green service innovation and firm performance. Lee and Song (2015) suggested that shipping knowledge can have a positive influence on creating logistics value only in the presence of ACAP. On the other hand, the importance of an individual's ability to acquire specific knowledge from the external environment is addressed in Ambulkar et al. (2016). Ambulkar et al. (2016) conceptualised that an individual's ACAP measured through a supply chain manager's risk mitigation competence and ability to absorb risk-related information has a positive influence on the success of risk mitigation programs. Thus, researches under this category predominantly focus on outcomes of developing knowledge-specific ACAP, but its antecedents are not well explained. However, Setia and Patel (2013) and Fayard et al. (2012) illustrated how technology especially IT and electronic integration can influence developing knowledge-specific ACAP. However, it can be argued that the dynamic nature of supply chain and environmental uncertainty could be one of the factors that motivate supply chain organisations to develop knowledge-specific ACAP. Further, it can be argued that, in the context of environmental knowledge, organisations that are uncertain about environmental outcomes tend to develop their capability to absorb environment and sustainability-related knowledge from external sources such as partners, competitors, customers, and industry.

2.6.5.4 Ambidextrous capability of absorptive and desorptive capacity

The internal capabilities of individual firms to absorb knowledge from the external environment are the main focus of discussion in most papers related to ACAP and SCM. However, it can be argued that, since the source of knowledge is external, it is necessary to equally focus the attention on capabilities of the source firm that can offer supportive mechanisms to allow the transfer of knowledge from source to recipient. If the source firm is incapable of developing mechanisms to aid effective transfer of knowledge, the return on the cost involved in developing the recipient firm's ACAP will not be satisfactory.

Only a few researchers have emphasised on developing the obligatory organisational capability to enable knowledge transfer in a dyadic relationship. Khan and Nicholson (2014) argued that "taking willingness to transfer knowledge as one side of a dualistic proposition, the contraposition would seem to be the ability to absorb knowledge that is transferred." From a social perspective, it can be argued that the willingness of the source firm to be conducive of knowledge transfer can be enhanced through building a close relationship with the source firm through mechanisms such as social interactions (Liu et al. 2017; Chowdhury et al. 2017). The current SCM literature has focused on investigating ways to develop the ability to absorb external knowledge and its influence on performance. However, it can be argued that in a dyadic relationship, firms should develop the ability to absorb as well as distribute knowledge to its partners whenever necessary.

Gosling et al. (2017) offers diverse views on supply chain learning, reviewed it as a process in which supply chain partners act together to learn from each other and resolve supply chain issues. In the process of supply chain learning, the role of the behavioural and learning ability of supply chain partners is paramount. From the capabilities perspective, it can be argued that a supply chain organisation with ambidextrous learning ability can acquire, assimilate, transform, exploit new external knowledge for value creation and distribute valuable internal information to its partners in the supply chain to encourage open innovation. In the context of intra-organisational learning, ambidexterity may be referred to as the dual orientation of learning, exploration, and exploitation. However, in the case of learning at the supply chain level, ambidexterity refers to the ability of firms to engage in internal learning processes such as exploration, assimilation and exploitation and, at the same time, teach external partners through similar processes of exploration and exploitation. Recent studies have discussed the importance of distributive capability (Whitehead et al. 2016) and desorptive capacity (Meinlschmidt et al. 2016) that can enable knowledge transfer in the buyer-supplier relationship. Both the terms ‘distributive capability’ and ‘desorptive capacity’ are used in the literature to refer to the ability to transfer knowledge from one entity to another. In Whitehead et al. (2016), distributive capability is conceptualised as an “accumulation of abilities associated with a source to transfer knowledge to a known recipient”. Emphasis on developing ambidextrous capability can be beneficial to both buyers and supplier, as it is vital to ensure bi-directional exchange of knowledge among members of supply chain, evading behavioural conflicts, and build supply chain relationship.

2.7 Operational performance

In this section, the literature review of operational performance dimensions is presented. The review particularly focuses on identifying the key dimensions of operational performance in the context of manufacturing.

2.7.1 Performance measurement

The growing competition in domestic and international markets is influencing firms to concentrate more on improving their performance to survive. Neely et al. (1995, p.1229) have defined performance measurement as “the process of quantifying the efficiency and effectiveness of an action”. Measuring firm performance could indicate how well firms do in achieving financial and marketing goals (Qrunfleh and Tarafdar, 2013). Since the main motive of stakeholders is to make a profit, various dimensions are used in the literature and practice to measure performance (Lii and Kuo, 2016)**.** Moreover, **“**performance measurement should be viewed as a context-dependent process” (Tseng and Liao 2015, p.86)**.** Also**,** Tseng and Liao (2015) contended that performance measurement can be viewed from two different perspectives: subjective and objective. The subjective perspective is concerned with measuring firm performance relative to a firm’s competitors. Whereas from an objective perspective, absolute measures are used as indicators for a firm’s performance. Neely et al. (1995) have stated that a firm’s operational performance measures the efficiency and effectiveness of its business activities.

Alttok (1997, p.1) perceived a manufacturing system “as an arrangement of tasks and processes, properly put together, to transform a selected group of raw materials and semi-finished products to a set of finished products". Manufacturing organisations rely on developing performance management systems to measure and monitor performances of the transformation process using various dimensions strategically determined by them. Kaydos (2000) cited H.James Harrington, "measurement is the first step that leads to control and eventually to improvement. If you can't measure something, you can't understand it. If you can't understand it, you can't control it. If you can't control it, you can't improve it". Performance management is a business process that determines the prosperity and growth of the manufacturing organisations (Bititci et al.1997). They further argued that information systems enable firms to manage performance and ensure that business actions are aligned to achieve the desired outcomes and create value for shareholders. Ideally, it is necessary for the organisations to strategically choose performance metrics and a performance management system to identify operational areas that require improvement (Neely et al. 1999; Pinheiro De Lima et al. 2013; Jain et al. 2011).

Consequently, to investigate the linkage between BDA capabilities and operational performance, the dimensions of operational performance that are widely used in the domain are assessed. Primarily, the SLR on BD studies (discussed in section 2.3) have shown that few studies have investigated the influence of analytics on various performance measures such as operational performance (Chae, Yang, Olson, et al. 2014; Chae, Yang and Olson, 2014; Jamehshooran et al. 2015a; Jamehshooran et al. 2015b), supply chain performance (Trkman et al. 2010; Oliveira et al. 2012), and supply chain agile performance (Sangari and Razmi, 2015). However, except Chen et al. (2015a), none of these studies has considered the impact of BDA capabilities on firm performance. Chen et al. (2015a) found that the use of BDA in SCM has a positive effect on business growth, measured through sales growth, market share, and expansion. Chen et al. (2015a)'s study is generalised based on the findings from a population representing all industry sectors around the world, but BDA practice is context specific. A large body of literature has investigated the relationship between Information Technology (IT) and operational performance. However, very little is known about the consequences of BDA, especially its impact on operational and innovation performance.

BDA is beneficial in terms of improving both financial and non-financial performance by means of cost reduction (Ge and Jackson 2014; Sonka 2014; Bock and Isik 2015), mass customization (Tien, 2012), improving logistics performance (Brandau and Tolujevs 2013), forecasting accuracy (Nita, 2015), recycling useable parts in manufacturing (Ge and Jackson 2014), improving quality (Sonka, 2014), on-time delivery (Yesudas et al. 2014), improved decision making (Hazen et al. 2014), increased visibility (Vera-Baquero et al. 2015; Schoenherr and Speier-Pero, 2015), real-time fault detection in manufacturing process (Aho, 2015), inventory planning, and network redesign (Tachizawa et al. 2015), better customer satisfaction, etc. Moreover, Laney (2015) analysed 40 different case studies and found that that the development of new products and services is the biggest benefit of BDA. For instance, Opresnik and Taisch (2015) argued that the potential of producing Big Data in manufacturing industries is high and this offers opportunities for creating services along with products (via servitization strategy), and consequently, organisations who utilize BDA can develop innovative products and service (P-S), improving existing P-S, and also can generate new revenue streams (via business models such as ‘Data resell’). The innovation performance of a firm is dependent on its knowledge creation activities (Bellamy et al. 2014). As mentioned in Dobrzykowski et al. (2015), the acquisition of critical information from the internal and external source is needed to improve innovation performance. BDA as an information processing tool improves the organisation’s information processing capabilities to innovate and gain a competitive advantage. Duan and Cao (2015) have investigated the impact of BA (descriptive, predictive and prescriptive analytics) on a firm’s innovation and found significant positive results. However, Duan and Cao (2015)’s study did not comprehensively investigate the phenomenon of BDA and their research focus is also not on SCM. So, either directly or indirectly, BDA has the potential to influence both the financial and non-financial performance of firms.

Table 2.15 Dimensions of operational performance from the literature review

Operational performance	**Cost reduction, order processing cost reduction with suppliers and customers**	**Cost**	**(Carter 2005), (Kim 2006) (Kim 2009), (Leuschner et al. 2013), (Mackelprang et al. 2014), (Yao et al. 2009), (Liu et al. 2012), (Ou et al. 2010) (Wu et al. 2014)**
	Inventory cost	Cost	(Glenn Richey Jr et al. 2009), (Li et al. 2014)

	Product cost	Cost	(Glenn Richey Jr et al. 2009), (Ağan et al. 2014), (Allred et al. 2011)
	Material and service cost	Cost	(Hollos et al. 2012)
	Labour cost	Cost	(Hollos et al. 2012)
	Production cost	Cost	(Hollos et al. 2012), (Li et al. 2014), (Gunday et al. 2011)
	Purchasing cost	Cost	(Allred et al. 2011)
	Supply chain cost	Cost	(Allred et al. 2011)
	Reduced operations cost	Cost	(Liao and Kuo 2014)
	Transportation costs	Cost	(Allred et al. 2011)
	Logistics cost	Cost	(Seo et al. 2014)
	Productivity, production flexibility	Flexibility	(Glenn Richey Jr et al. 2009), (Ng et al. 2015), (Gunday et al. 2011)
	Supplier performance		(Carter 2005)
	Response time for product design changes	Flexibility	(Kim 2006), (Kim 2009)
	Response to changing demand, unexpected challenges.	Flexibility	(Liu et al. 2013b), (Glenn Richey Jr et al. 2009), (Liu et al. 2013a), (Yu 2015) (Li et al. 2014), (Bruque-Camara et al. 2016)
	Response time for product volume changes	Flexibility	(Kim 2006), (Kim 2009) (Bruque-Camara et al. 2016)
	Responsiveness to special delivery requests	Flexibility	(Miller et al. 2013)
	Lead time to customers	Flexibility	(Hallavo 2015)
	Accuracy of order processing for customers, percentage of late or changes deliveries	Quality	(Kim 2006), (Kim 2009), (Miller et al. 2013)
	Perceived product quality by customers, firm image	Quality	(Singh and Power 2009), (Singh and Power 2014), (Grekova et al. 2015)

	Rapid confirmation of orders	Flexibility	(Hefu Liu et al. 2013b)
	Speed, Speed of order handling	Flexibility	(Kim 2006), (Kim 2009) (Golicic and Smith 2013)
	Response time for product returns or after-service	Flexibility	(Kim 2006), (Kim 2009)
	Reduced return	Quality	(Khan K et al. 2009)
	Quality, Product quality	Quality	(Hsu et al. 2008), (Singh and Power 2009), (Golicic and Smith 2013),(Leuschner et al. 2013), (Mackelprang et al. 2014), (Wu and Chuang 2010), (Glenn Richey Jr et al. 2009), (Huo et al. 2014), (Yu 2015), (Ağan et al. 2014), (Liu et al. 2012), (Tan et al. 2010), (Narasimhan et al. 2008), (Ou et al. 2010), (Ng et al. 2015), (Hollos et al. 2012), (Allred et al. 2011), (Grekova et al. 2015), (Gunday et al. 2011)
	Delivery time, delivery cycle time, short lead time for fulfilling orders, reliable delivery	Time	(Kroes and Ghosh 2010), (Leuschner et al. 2013), (Mackelprang et al. 2014), (Hefu Liu et al. 2013b), (Hefu Liu et al. 2013a), (Gligor and Holcomb 2012), (Singh and Power 2014), (Kirchoff et al. 2016), (Yu 2015), (Liu et al. 2012), (Hollos et al. 2012), (Seo et al. 2014), (Miller et al. 2013), (Bruque-Camara et al. 2016)
	Manufacturing cycle time, reduced make time, short production cycle time	Time	(Kroes and Ghosh 2010), (Khan K et al. 2009), (Ou et al. 2010)
	On-time delivery	Flexibility	(Kroes and Ghosh 2010), (Glenn Richey Jr et al. 2009), (Khan K et al. 2009), (Ou et al. 2010), (Allred et al. 2011), (Li et

			al. 2014), (Hallavo 2015), (Bruque-Camara et al. 2016)
	Processing costs	Costs	(Kroes and Ghosh 2010)
	Supply chain costs	Costs	(Klassen and Vereecke 2012)
	Inventory levels, inventory turnover	Efficiency	(Singh and Power 2009), (Singh and Power 2014), (Kirchoff et al. 2016), (Yao et al. 2009) (Seo et al. 2014), (Li et al. 2014), (Hallavo 2015)
	Inventory visibility	Flexibility	(Seo et al. 2014)
	Product availability in stock	Flexibility	(Hallavo 2015)
	Time taken for new product development, time for introducing new products or services	Time	(Singh and Power 2009), (Hefu Liu et al. 2013a), (Singh and Power 2014)
	Delivery performance, delivery speed, accuracy	Delivery	(Singh and Power 2009), (Huo et al. 2014), (Yu 2015), (Seo et al. 2014), (Li et al. 2014) (Gunday et al. 2011)
	Flexibility	Flexibility	(Golicic and Smith 2013), (Leuschner et al. 2013), (Mackelprang et al. 2014), (Wu and Chuang 2010), (Liu et al. 2012), (Bruque-Camara et al. 2016)
	Flexibility in volume	Flexibility	(Huo et al. 2014)
	Product mix flexibility	Flexibility	(Huo et al. 2014)
	New product flexibility, New product development capability	Flexibility	(Huo et al. 2014), (Allred et al. 2011), (Bruque-Camara et al. 2016)
	Innovation	Innovation	(Leuschner et al. 2013), (Hollos et al. 2012), (Gunday et al. 2011)
	Product innovation lead times	Innovation	(Glenn Richey Jr et al. 2009)
	Overall performance	Efficiency	(Cho et al. 2008)

Environmental/ social performance	Decrease energy cost	Cost	(Grekova et al. 2015), (Zailani et al. 2015)
	Water and waste water cost	Cost	(Grekova et al. 2015)
	Water and waste water treatment cost	Cost	(Grekova et al. 2015)
	Decreased CO_2 emission	Efficiency	(Laari et al. 2015)
	Waste reduction	Efficiency	(Laari et al. 2015)
	Energy consumption	Efficiency	(Laari et al. 2015)
	Dealing with environmental issues	Efficiency	(Laari et al. 2015)
	Decrease in consumption of hazardous materials	Efficiency	(Laari et al. 2015)

From Table 2.15 it appears that the recent literature portrays a divergent view on the operational performance dimensions and it has been the main debate for several years. However, it seems that quality, cost, time (speed and delivery) and manufacturing flexibility have received significant attention while performing manufacturing tasks and it is well articulated in the literature. As Peng et al. (2008) observed, perhaps the measures of operational performance are multidimensional in nature. Thus, considering the indications from various literature, this study will measure operational performance of the UK manufacturing companies by means of four dimensions such as quality, cost, manufacturing flexibility and time, as these are the key competitive priorities of manufacturing organisations (Porter 2009; Slack et al. 2007).

2.7.2 Quality

Quality is one of the important performance measures that can indicate a manufacturer’s competitiveness (Nada et al. 2006). Quality is unquestionably a major concern to manufacturing organisations (Slack et al. 2016). In order to successfully compete in today’s globally competitive market, organisations should understand the importance of quality as it is one of the determining factors of market share (Calantone and Knight 2000). The organisation's ability to achieve good quality or the lack of quality may affect the entire supply chain, and it is an essential precondition to achieving operational excellence. As verified by Ferdows and De Meyer (1990), quality may influence other performance outcomes such as cost, flexibility, and dependability. The two major elements included in the quality performance dimension are the quality of the product or services, and the process quality that involves delivery of the product or services (Porter 2009). “Quality robust products are products that can be produced uniformly and consistently in adverse manufacturing and environmental conditions” (Heizer et al. 2017, p.224).

There are various definitions of the term ‘quality’ available in the literature. According to Garvin (1984, p.40) quality is a “slippery concept, easy to visualize and yet exasperatingly difficult to define”. Similarly, Shetty (1987) contended that quality is a complex concept. Moreover, Shetty (1987) mentioned that better quality products reduce cost by means of reducing scrap, rework, reducing work in process, better utilisation of tools and reduced warranty claims. Calantone and Knight (2000) have defined quality from the perspective of product characteristics, and when a product’s features and performance needs exceed the expectation of customers, then it can be considered as a high-quality product. Garvin (1984) further mentioned the key dimensions that can be used to assess product quality: 1) performance, 2) features, 3) reliability, 4) conformance, 5) durability, 6) serviceability, 7) aesthetics, and 8) perceived quality. As argued by Shetty (1987), the suitability of these dimensions of quality may differ extensively between firms in the same industry. The choice of the quality dimension depends on both the manufacturers’ perceptions of the customers’ expectations and their capability to produce products that can satisfy certain quality demands. Consequently, a firm that wants to strategically compete on the basis of product quality, needs to understand customers’ expectations accurately and incorporate them into the product design and development process. In order to ensure that the manufactured products conform to the quality requirements, the quality characteristics or standards must by predetermined and aptly monitored. Slack et al. (2016) prescribed that maintaining quality requirements should start from the design stage. Similarly, Fynes and De Búrca (2005) highlighted the significance of the design stage on the development of quality products and argued that ‘‘quality is designed into the product at least as much as it is built in during manufacture” (P.1). Further, Nada et al. (2006) also contended that product quality is influenced by both the product design and the degree of conformance. In the product design phase, practices such as

Total Quality Management (TQM) influence the product design and ultimately the product quality.

However, some researchers argue that apart from the design stage quality is also affected by the configuration of processes producing the product. For instance, Inman et al. (2003), illustrated various scenarios of automotive companies such as Ford, Jaguar, and Toyota where the formation and design of the manufacturing system greatly affect the product quality. So, the quality of the manufacturing system has to be assessed at the early stages to produce a high-quality product at a reduced cost. Moreover, since the manufacturing environment is changing at a fast rate along with fluctuating demand, an efficient manufacturing system that can be reconfigured rapidly can influence the development of high-quality products. Kaynak (2003) asserted that the process perspective of quality is best met by the preventive method of ensuring product quality. Reducing the variations in manufacturing processes leads to increased product quality (Yeung et al. 2005; Kaynak 2003; Forza and Filippini 1998).

Moreover, product quality enhances the firm's reputation via increased customer satisfaction and also increases productivity and reduces cost (Calantone and Knight 2000). Product quality is empirically found to be an important criterion to enhance customer satisfaction (Fornell et al. 1996). Moreover, several studies have found a positive link between product quality and competitiveness (Phan et al. 2011), reputation (Gjerde and Slotnick 2004), increased market share and profitability (Calantone and Knight 2000) and return on investment (Buzzell 2004). So, quality indicates the manufacturing firm's performance in terms of how effective and efficient it is in designing, producing and delivering products to its customers.

2.7.2 Cost

Since the start of globalisation, cost has become an important competitive strategy and firms actively identify key costs and the drivers of it (Blocher et al. 2010). This is because, with the increase in the intensity of competition, the efficiency of production is determined by cost performance. For organisations that intend to compete based on price, it is vital to decrease the cost of manufacturing compared to their rivals. If the price is lower than their competitors, companies can either generate more profit or increase their market share (Porter 2009). Moreover, the cost-based strategy can be an important tool for targeting products to a niche market and to deter competitors and new entrants from entering the market. Porter's competitive strategy includes cost leadership and differentiation strategies (Figure 2.29). Porter (1985) defined cost leadership as "a firm sets out to become the low-cost producer in its industry" (p.12). Low cost, denoting efficiency, remains as the strategic intent of most organisations. Whereas, firms adopting a differentiation strategy "seek to be unique in their industry along with some dimensions that are widely valued by buyers" (Porter 1985, p.14). Unlike cost leaders, firms intending to differentiate choose an attribute that is different from their competitors and aim at cost parity with its rivals. Scholars have identified ways to implement cost leadership strategies and differentiation strategies (Hill et al. 2015a; Brennan 2011; Blocher et al. 2010). According to Blocher et al. (2010), the execution of a cost leadership strategy requires tight control over cost, periodic control reports, well-structured policies and operations strategy and incentives for meeting targets specific to cost. Whereas, for the execution of a differentiation strategy, the organisation must build strong coordination among key business functions such as research & development, product design, manufacturing and marketing. Organisations adopt a competitive strategy depending on their market position, strategic goals and organisational culture. Still, managing cost is one of the key priorities for manufacturing businesses to sustain in the competitive environment.

Moreover, recognising different types of costs is essential to develop cost management techniques. There are several types of costs commonly used in literature such as material costs, overhead cost, labour costs, quality costs, transportation costs, etc. (Amoako-Gyampah and Boye 2001). Omachonu et al. (2004), by utilising the quality cost model, found out that the increase in costs such as the appraisal and prevention cost increases the product quality. Material related quality cost indicates the cost of raw material inspections, scrap and rework, supplier evaluation cost, etc. Machine-related quality cost indicates the costs involved in preventive maintenance, repair, rework, calibration of machines, etc. A brief overview of quality costs is provided in Table 2.16. Brennan (2011) argued that manufacturing organisations should consider both the short-term and long-term costs of manufacturing a product. For instance, a manufacturer could utilise cheap materials with an intention to save direct costs but may face high repair and warranty costs.

		Competitive advantage	
		Lower cost	Differentiation
Competitive Scope	Broad Target	1. Cost Leadership	2.Differentiation
	Narrow Target	3A. Cost Focus	3B. Differentiation Focus

Figure 2.29 Porter's generic competitive strategy

Source: (Porter 1985)

Table 2.16 Taxonomy of quality costs

Prevention Costs	Appraisal Costs	Internal Costs	External Costs
• Employee training. Process capability studies. • Surveys of vendors, suppliers, and subcontractors.	• Inspection and testing of products. • Capital cost and ongoing maintenance of inspection and testing equipment. • Cost to process and report inspection data. • Design reviews. Expense reviews.	• Scrap and rework. • Charges related to late payments. • Inventory costs to allow for defective products. • Engineering change costs for design correction. • Premature failure of products. • Correction of documentation.	• Warranty repairs. • Field service personnel training. • Complaint handling. • Customer dissatisfaction. • Future business losses Litigation.

Source: (Brennan 2011, p.61)

According to Blocher et al. (2010), organisations manage costs by collecting both financial and non-financial information. Financial information includes incurred costs and revenues generated by the organisation. Non-financial information comprises specifics about productivity and quality. Thus, cost management is “the development and use of cost management information” (Blocher et al. 2010, p.3). Organisations create cost management systems focusing on internal and external cost reporting by tracking its operational data. Management of cost appears to be a driving force for all strategic choices, especially for operational strategy. Firms that face the issue of increasing business cost may be less inclined to develop plans that may lead to flexibility (Amoako-Gyampah and Acquaah 2008). So, firms that encounter an increased level of cost rely on cost related information to create systems that can monitor and control by developing appropriate strategies. Blocher et al. (2010) have illustrated the importance of cost management via Procter & Gamble (P&G). As one of the leading fast-moving consumer goods manufacturers, P&G places emphasis on cost reduction strategies and utilises process and product simplification approaches. Since the company has more than 50 brands of different size packages and flavours, they faced high complexity resulting in increased costs related to manufacturing, inventory, administration, sales and distribution and operation costs. However, after the implementation of the cost reduction strategy, they reduced the number of product varieties and as a result, the company witnessed the decrease of supply chain costs and the increase of product quality and innovation. Similarly, manufacturing firms can deliver superior value to customers via cost-based competitive strategies (Wagner 2006). These can be achieved either via offering a product with better value at a higher price or at a lower price for the same level of benefits as competitors. To proceed with a lower price option, firms need to come up with innovative cost reduction strategies. Similarly, previous studies such as Shehab and Abdalla (2001) have estimated that the product design phase constitutes 6% of the development cost but determines 70% of production cost, and the variable cost is estimated to reduce 22% using flexible production volume. Thus, the effectiveness of managing various costs can be a useful indicator of firm performance as it could significantly increase competitiveness and sustainability.

2.7.3 Flexibility

Manufacturing flexibility is truly a multi-dimensional concept generally known as a critical component to achieving competitive advantage (Jain et al. 2013; Gupta and Somers 1992). Flexibility is one of the properties highly desired by manufacturing firms as it has become a requirement for survival under environmental uncertainty (Shi and Daniels 2003; Patel et al. 2012; Sethi and Sethi 1990; Merschmann and Thonemann 2011; Scherrer-Rathje et al. 2014; Wall 2003). Flexibility increases the responsiveness of manufacturing systems and supports the utilisation of the available resources to the fullest. Also, it is widely recognised that flexibility can improve operational performance and enhance the ability of a manufacturing firm to adapt to internal and external disturbances (Sethi and Sethi 1990; Jain et al. 2013). For instance, Francas et al. (2011) illustrated how improving flexibility has helped Renault’s manufacturing systems to become more agile and adaptable. Renault’s strategy focuses on changing an inflexible one-plant polices to a highly flexible system capable of manufacturing multiple products. This strategy is preferred by the organisation to overcome the unpredictable demand. Similarly, due to improved flexibility, BMW's US manufacturing plant avoided the layoffs during the financial crisis in 2008 and also evaded the negative impact of the economic crisis on sales (Rogers et al. 2011). Moreover, flexibility is found to be an important element for improving innovation (Oke 2013a).

Since it is a multi-dimensional concept, there is no one definition generally accepted (Gerwin 1993; Barad and Sipper 1988). Barad and Sipper (1988, p.237) defined flexibility as "the capability to bend without breaking, to be adaptable". Apart from these, various definitions of flexibility frequently used by researchers are available in the literature (Fredericks 2005; Jain et al. 2013). A selection of important definitions from Jain et al. (2013) is provided in Table 2.17. Most definitions available in the literature relate flexibility to uncertainty management. Gerwin (1993) identified different types of uncertainties such as market that require different kinds of products, shorter product life cycle, product characteristics, fluctuating product demand and downtime of machines. Depending upon the type of uncertainties, organisations must develop specific strategies and adaptive methods to become flexible and reduce uncertainties. For instance, in the case of uncertainties due to the length of the product life cycle, it is suggested that firms should use life extension practices and develop changeover flexibility, with the "ability to quickly substitute new products for those currently being offered" (Gerwin 1993, p.399). Moreover, there are various dimensions of flexibility discussed in the academic literature. D'Souza and Williams (2000) categorised flexibility into two externally driven and internally driven dimensions as given in Table 2.18. Apart from dimensions such as volume, variety, process and material flexibility, there are various types of flexibility (such as machine flexibility, transfer flexibility, expansion flexibility, market flexibility and labour flexibility) elaborately discussed in the literature (Barad and Sipper 1988; Jain et al. 2013).

Table 2.17 Definitions of manufacturing flexibility

References	Definitions of manufacturing flexibility
(Nagarur 1992)	"The ability of the system to quickly adjust to any change in relevant factors like product, process, loads and machine failure".

(Koste and Malhotra 1999)	"The ability of manufacturing function to react to changes in its environment without significant sacrifices to firm performance".
(Watts et al. 1993)	"The ability to implement changes in the internal operating environment in a timely manner at a reasonable cost in response to changes in market conditions".
(Gupta and Goyal 1989)	"The ability of a manufacturing system to cope with changing circumstances or instability caused by the environment"

Source: (Jain et al. 2013)

Table 2.18 Dimensions of manufacturing flexibility

Category 1: Externally-driven flexibility dimensions

Volume flexibility	This dimension of flexibility represents the ability to change the level of output of a manufacturing process
Variety flexibility	This dimension represents the ability of the manufacturing system to produce a number of different products and to introduce new products. Researchers have suggested the use of product mix and product modification as components of this dimension of manufacturing flexibility
Category 2: Internally-driven flexibility dimensions	
Process flexibility	This dimension represents the ability of the system to adjust to and accommodate changes/disruptions in the manufacturing process. Examples of these changes / disruptions found in the literature are, machine breakdowns, changes in the production schedules, or job sequencing
Materials handling flexibility	This dimension represents the ability of the materials handling process to effectively deliver materials to the appropriate stages of the manufacturing process and position the part or the material in such a manner as to permit value adding operations

Source: (D'Souza and Williams 2000)

Moreover, Sánchez and Pérez (2005) investigated the importance of flexibility and developed a hierarchical framework to delineate the dimensions of flexibility. As shown in Figure 2.30, the dimensions of flexibility can be categorised into basic, systems, and aggregate level. The sub-dimensions of basic-the shop floor flexibility includes product, volume and routing. Since the focus of this research is on the operational performance of a manufacturing plant, the flexibility dimensions of shop floor level are discussed further. Accordingly, product flexibility denotes the ability of the firms to handle unexpected orders and the special requirements of customers and to produce products of diverse characteristics such as varying sizes, colours, etc. Moreover, the authors argued that to develop product flexibility, collaboration with other functional departments such as marketing, design and engineering is a precondition. Volume flexibility is defined as the "ability to effectively increase or decrease aggregate production in response to customer demand" (Sánchez and Pérez 2005, p.685). Similarly, D'Souza and Williams (2000, p.580) defined volume flexibility as "the ability to change the level of the output of a manufacturing process". Barad and Sipper (1988, p.239) defined it as "ability to operate a flexible manufacturing system profitably at different production volumes". Whereas, routing flexibility is defined as "the capability of processing a part through varying routes by using alternative machines, flexible material handling, and flexible transporting network".

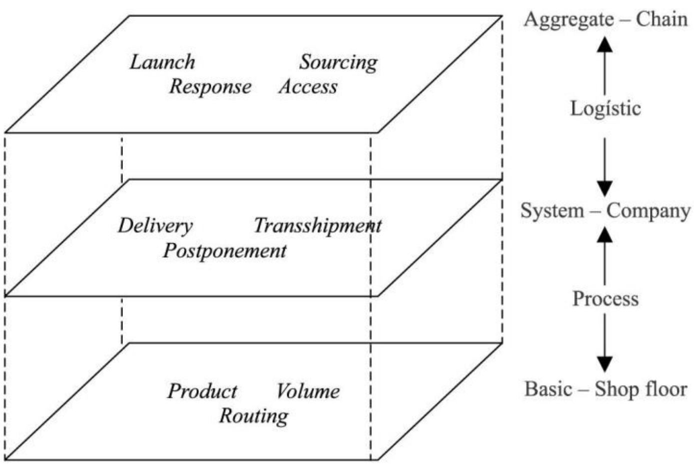


Figure 2.30 Dimensions of flexibility

Considering the importance of flexibility, a large number of researchers have investigated the identification of its antecedents. Rosenzweig et al. (2003) found out that information sharing enhances flexibility, which, in turn, improves customer satisfaction and sales growth. Chan and Chan (2009) have found out that information exchange can improve coordination and enhance flexibility. Fredericks (2005) argued that flexible utilisation and coordination of resources such as information technology is a way to manage uncertainty in the business environment. Ozer (2002) identified acquiring and using flexible hardware and software technology as a source of flexibility. Golden and Powell (1999) put forward that improving flexibility is contingent upon the degree of information shared via inter-organisational systems. Similarly, web services, which enable data acquisition, are argued to enhance flexibility (White et al. 2005). Fredriksson and Gadde (2005) illustrated that data sharing is an important antecedent of flexibility in the case of Volvo's car manufacturing system. Improvement in flexibility could increase transparency, reduce inventories, reduce remanufacturing, overproduction and enhances trust (Rosenzweig et al. 2003). Thus, an organisation's ability to flexibly adapt their manufacturing system to improve operational efficiency and effectiveness is a useful indicator of operational performance.

2.7.4 Time-based performance

Organisations increasingly focus on time-based competition as a strategy for delivering value to customers (Blackburn 1992). Koufterosa et al. (1998) argued that manufacturing firms are experiencing a paradigm shift. Traditional manufacturing systems based on the industrialisation era are driven by efficiency, but due to increased customer demand and competition, the post-industrial systems are driven by the quick response. The emergence of 'time-based competition' is traced back to the efforts of George Stalk who devised the term emphasising the importance of time for creating a sustainable competitive advantage (De Toni and Meneghetti 2000). Stalk Jr. and Hout (1990) argued that recently customers have specific demands of 'what product/service they want, when they want it, and where they want it to deliver'. These changes in the consumer behaviour require firms to be more agile, and speed/time must be considered as an important component. Accordingly, from the 1990s, time-based competition is acknowledged by the academics and practitioners as a key success factor for competitiveness (So and Song 1998; Blackburn 1991; Stalk Jr. and Hout 1990). Similarly, Koufterosa et al. (1998) and Kim and Tang (1997) have argued that during the early 1980s high quality and low cost are considered as a dominant indicator of operations performance. However, nowadays, manufacturers perceive time-based competition as a fundamental source of sustainable competitive advantage. Blackburn (1992) called time-based competition as the "next battleground". Stonich (1990) perceived this paradigm change as "the next strategic frontier".

Moreover, Blackburn (1991) suggested that the origin of time-based competition is closely connected to the implementation of Just-In-Time(JIT) manufacturing practices. Manufacturers who are the early adopters of JIT practices are the pioneers of time-based competition. Manufacturers who implemented JIT recognised the potential of time-based manufacturing to significantly reduce the response time (Tu et al. 2006). Further, it is elaborated that JIT and time-based manufacturing are underpinned by the same philosophy but with a dissimilar focus. While JIT practices are internally oriented with an objective of reducing waste and non-value-adding activities, time-based manufacturing is externally focussed with an objective of reducing throughput time and rapid response to consumer demand. Ultimately, both the approaches intend to create a manufacturing system that can achieve significant cost reduction and improved customer satisfaction.

Koufterosa et al. (1998) developed a framework and defined a set of seven practices that constitute time-based manufacturing. The seven factors of time-based manufacturing practices are: 1. “shop floor employee involvement in problem-solving”, 2. “re-engineering set-up”, 3. “Cellular manufacturing”, 4. “Quality improvement efforts”, 5. “Preventive maintenance”, 6. “Dependable suppliers”, 7, “pull production”. Briefly, re-engineering set-up practices indicate the manufacturing systems ability to reduce set-up time; cellular manufacturing practices indicate the ability of the system to produce units in a product-oriented layout; practice of employee involvement in problem-solving indicate firms ability to make time-based changes; quality improvement practices indicate the ability of the system to reduce waste and increase quality; preventive maintenance practices indicate the degree to which the machine defects are proactively monitored and routinely maintained; practices that create dependable suppliers indicate the ability of the system to reduce throughput time and on-time delivery as suppliers performance can improve service quality; and production pull indicates firms ability to timely meet the demand pull from the customers and reduce waiting time. Consequently, it is argued that the implementation of these practices can increase responsiveness by minimising the time from every stage of the value-delivery systems. Scholars have found out that higher productivity, better customer service and increased market share are some of the outcomes of efforts put in place to redesign the manufacturing processes to compress time (Blackburn 1991; Nahm 2003; Stalk Jr. and Hout 1990), and sustainable competitive advantage is the prize for achieving speed in value-delivery systems (Koufterosa et al. 1998).

Phase \ Time performance	Internal	External
Product development	Time-To-Market (TTM)	(Frequency of introducing) -new products -existing product improvements
Procurement Production Distribution	Lead Time (LT)	Delivery Time(DT) -Speed -Punctuality

Source:(De Toni and Meneghetti 2000)

Figure 2.31 The dimensions of time performance

Further, Droge et al. (2004) argued that some manufacturing firms aim to improve the speed of product development, but others may focus on improving speed on various facets such as manufacturing, procurement and distribution. As shown in Figure 2.31 the time-based performances can be categorised into internal and external perspective. Lead time and delivery time are the dimensions that relate to the manufacturing process. Whereas, time-to-market and frequency of new product development and market are the time dimensions related to product development. According to Little (1961), there is a relationship between inventory and time. Little’s law proves that time-in-process is the product of work-in-process and expected time between successive arrivals ($L=\lambda W$). The increasing overhead cost is due to a large number of products lingering in the manufacturing system for a prolonged time (De Toni and Meneghetti 2000). Reducing lead time can decrease work-in-process, which, in turn, can decrease working capital and offer greater inventory turnover. Koufterosa et al. (1998) acknowledged the claim of Little (1961) and argued that reduced inventories by faster set-up time could improve cost performance and minimise risks.

Thus, time in a shop floor environment is an important factor that can influence the choice of customers when choosing an organisation (Slack et al. 2016). In recent times, customers consider the total cycle time of product/service from start to finish stage as an increasingly important aspect for satisfaction (Blackburn 1992). For instance, the internal process speed of purchase and make the function of manufacturing will influence the delivery time. Also, the operation time which is the difference between the time a customer makes a purchase request for a product/service and the time of delivering it, on competitive grounds is more significant in reducing costs and improving customer satisfaction (Slack et al. 2007). Urban (2009) also opined that the on-time delivery of products and services is inevitable to sustain competition. Spanner et al. (1993) observed that time-based practices enable manufacturers to establish strategic alliances with their supply chain partners and build commitments to produce high-quality products. In the case of poor quality products, extra time is required to make corrections and improve quality. Time/speed is an important manufacturing strategy to increase productivity (Nahm 2003; Schmenner et al. 2009; Stalk Jr. and Hout 1990). Hence, time is considered in this study as a significant indicator of the operational performance of manufacturing firms.

2.8 Innovation performance

The term 'innovation' is widely used in business practice and academic literature, but it lacks a consensual definition in the management literature (Sattler 2011). Christiansen (2000) compares innovation to a complex manufacturing process as it involves several steps and the performance level of each step determines the quality of the innovation output. However, unlike manufacturing processes, innovation processes are not repeated and so customisation is required of the definition. According to Maital and Seshadri (2007, p.29), "innovation occurs when an invention, related to a product, service or process in some part of the organisation's value chain, is joined with a business design, which in turn is implemented with discipline and skill through innovation management". The link between invention, innovation and competitive advantage is depicted in Figure 2.32. Davenport et al. (2006) argued that leading companies in the world are transforming themselves from the industrial economy to the innovation economy. Organisations are finding new ways to compete and take advantage of their innovation potential. A metaphor ($Energy_{business}$ = Management × (Innovation × Speed)) based on the classic equation of Einstein's $E=MC^2$ is used to understand the characteristics of innovation economy (Davenport et al. 2006). This represents that the energy of the business model is equivalent to the product of management strategy orientation concerning innovation and speed. Moreover, Sundbo and Fuglsang (2002) have discoursed the nature of innovation using the following statement:

“Innovation is seen as a complex, anarchistic and unsure social process containing many alternative possibilities. Innovation is a game where the players become occupied with defining strategies and roles for themselves and beat other players. They are creative players, thus they invent not only new clever moves, but also new, or differing, rules of the game.” (p.7).

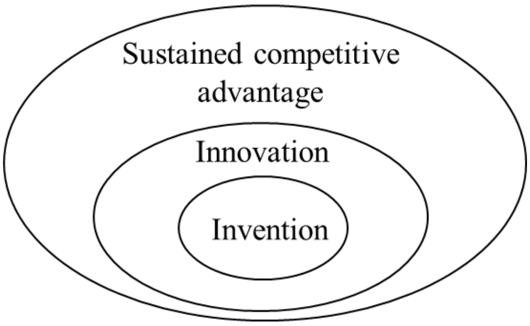


Figure 2.32 Invention, innovation and sustained competitive advantage

Source: (Maital and Seshadri 2007)

The domain of innovation is broad and several scholars have performed a review of innovation literature (Adams et al. 2006; Pittaway et al. 2004; Wolfe 1994; Crossan and Apaydin 2010; Gopalakrishnan and Damanpour 1997). An excerpt of the findings from these literature reviews is used in this research to delineate the concept of innovation, innovation process, innovation types and its effect on competitive advantage. There are several definitions of innovation available in the literature. In the seminal work of Thompson (1965, p.2), innovation is defined as “generation, acceptance, and implementation of new ideas, processes, products or services”. Maital and Seshadri (2007) defined innovation as “the practical refinement and development of an original invention into a usable technique or product;’ or, a process in which creativity is applied to every facet of an organisation’s value chain, from beginning to end, to develop new and better ways of creating value for customers” (p.29). Crossan and Apaydin (2010, p.1155) defined innovation is “production or adoption, assimilation, and exploitation of a value-added novelty in economic and social spheres; renewal and enlargement of products, services, and markets; development of new methods of production; and the establishment of new management systems. It is both a process and an outcome”. Moreover, Crossan and Apaydin (2010) argued that innovation diffusion “refers to a process taking place after innovation” and it is not related to innovation itself.

Gopalakrishnan and Damanpour (1997) have argued that researchers have measured innovation performance at various levels: 1. Industry level, 2. Organisational level, 3. Subunit level, 4. Innovation level. At the organisational level, most studies were aimed at understanding the characteristics of innovating and non-innovating organisations and describe the consequence of the innovation process. At the industry level, studies have focused on factors that influence innovation development, the observable pattern of innovation and industry level expenditure on research and development (R&D). At the sub-unit level, it is found out that the main focus of the scholars is on the R&D department and the antecedents of innovation. It is identified that several factors such as communication, decision making, tenure and diversity of R&D teams and leadership have played a significant role in improving innovation. On the other hand, studies on innovation level analysis focus on innovation characteristics (such as innovation types and innovation complexity) and the diffusion of an innovation within an industry or organisation.

Sundbo and Fuglsang (2002) have discussed the 5 stages of the innovation process: 1. the collection of information and ideas on a problem, 2, researching, 3. conception and development (C&D), 4. production of the solution, 5. marketing of the solution. In the first stage, the scientific and technical information from the firm's internal and external environment has to be collected as it is the main source of innovation. The second stage of the innovation process involves the activities that can combine old stocks of knowledge to create new knowledge. In the third stage, tests and experimentations are the common activities that are aimed at the transformation of ideas into a solution. In stage four, the solution is converted to an actual product or services. The final stage is both internally and externally oriented, as the new solution is marketed to the internal and external clients of the organisation.

Gopalakrishnan and Damanpour (1997) categorised the types of innovation into three sets composed of six different types of innovations. The innovation of product and process reflects the outcomes of the innovation process. Radical and incremental innovation refers to the degree of change involved in the adoption of innovation. Radical innovation involves the incorporation of new knowledge to a large degree but may bring uncertainties and risks, and requires a change in existing practices (Song and Thieme 2009). Whereas, incremental innovation is characterised by minor changes to existing practices and the level of uncertainty and risks are relatively low (Figure 2.33). Similarly, Jansen et al. (2006) have categorised innovation into exploratory and exploitative innovation. Exploratory innovation is similar to radical innovation but has a specific aim of meeting a new customer's or market's demands. Inversely, exploitative innovation is alike incremental innovation intended to satisfy the needs of current customers and markets. Lastly, technical and administrative innovations reflect the distinction between technology and social structures. While the administrative innovations related to innovations in organisational structures, human resources and management process, technical innovations are more product, processes and technology oriented.

Type of innovation

Task of innovation	Incremental	Radical
Predesign	• Moderate levels of uncertainty and idiosyncratic transactions. • Supplier involvement is moderately beneficial	• Highest levels of uncertainty and idiosyncratic transactions. • Supplier involvement is least beneficial.
Commercialization	• Lowest levels of uncertainty and idiosyncratic transactions. • Supplier involvement is most beneficial.	• Moderate levels of uncertainty and idiosyncratic transactions. • Supplier involvement is moderately beneficial.

Figure 2.33 Radical vs Incremental innovation

Source: (Song and Thieme 2009)

In an attempt to explain the antecedents of innovation, Hurley and Hult (1998) have found out how the organisational culture and learning orientation affects organisational innovativeness. Similarly, Hurley and Hult (1998) argued that "organisational learning is synonymous with the capacity to innovate." (p.46). Joshi and Sharma (2004) empirically verified that development of customer knowledge is the antecedent of new product development, which implies the strategic significance of customer knowledge for innovation performance of manufacturing firms. Thompson (1965, p.11) stated that "the innovative organisation will allow that diversity of inputs needed for the creative generation of ideas." Similarly, Mauerhoefer et al. (2017) have empirically found out that the use of information and communication technologies have a greater influence on improving the process of new product development. Hence, for improving innovation, organisations require diverse sources of knowledge. BDA can be considered as a source of obtaining knowledge from diverse sources required for generating creative ideas.

2.9 Summary of literature review and Research Gaps

In summary, this chapter has covered a comprehensive literature review of various concepts related to this research. The adoption of the systematic literature review methodology allowed the researcher to quantitatively summarise the existing contributions and identify several potential research gaps. The next section presents the research gaps addressed in this study and the rationale behind them.

2.9.1 Research Gaps

The systematic literature analysis highlights theoretical and empirical gaps.

First, the SLR of BDA in SCM revealed that research in this domain is emerging. Compared to other fields of research, BDA in the SCM domain is still in a nascent stage. Of the limited prior research, most of the studies are conceptual in nature (refer Page 25, Section 2.3.2.6, Figure 2.9), which imposes the need for both quantitative and qualitative studies. The BDA capabilities maturity (BDAM) construct is under-researched and its measures are not well established. In particular, the researcher did not find any empirical study relating to BDAM in the supply chain context. So, there is a need to develop measures for the dimensions of BDA Maturity (BDAM).

Second, another significant gap in the literature is that there is no empirical evidence illustrating the relationship between intra-organisational BDAM, Absorptive Capacity, supply chain analytics, data and information quality, and operational and innovation performance. Only a few survey-based studies exist in the domain of BDA and SCM (refer Page 25, Section 2.3.2.6, Figure 2.9). These empirical studies assess the role Business Analytics or Business Intelligence in SCM and its influence on SCP or OP (Trkman et al. 2010; Chae, Yang and Olson 2014; Kwon et al. 2014; Jamehshooran et al. 2015b; Chen et al. 2015a), but not in the context of BDA maturity. Also, its impact on operational and innovation performance is not well researched in recent studies (Dubey et al. 2015; Wamba et al. 2017; Ji-fan Ren et al. 2017; Gupta and George 2016). Moreover, considering the context of the previous research studies, only Cao et al. (2015) have focused on the UK context. However, they do not meticulously investigate the relationship specific to BDA and manufacturing supply chains in the UK. In addition, the underlying mechanism through which the practice of BDA can attain maximum value to the organisation needs comprehensive investigation.

Third, the empirical studies that exist have mainly focused on large organisations and little is known about the UK SMEs ability to possess BDA capabilities. As discussed in section 1.3, SMEs are arguably deficient of BDA capabilities and the concept of the digital divide caused by BDA has not investigated in the previous studies (Hilbert 2016; Arunachalam et al. 2018; Andrejevic 2014). So, this research also sets an agenda to empirically investigate the BDA capabilities of SMEs and examine the concept of the digital divide by using the maturity model. Therefore, this research aims to fulfil these knowledge gaps and make a significant contribution to the academic community and business practice. Consequently, to empirically investigate and address these research gaps a conceptual framework is developed, which is discussed in the following sections. The proportion of SMEs that have utilised big data analytics is found to be negligible (SAS 2013). The disparity in the adoption of these innovative technologies would certainly hinder the growth of SMEs. Perhaps, on the capabilities framework proposed in this research, SMEs would fall under the category of Data Poor and Information Poor. It is important to investigate the effect of BDA adoption on the extension of the digital divide between SMEs and large organisations.

The antecedents of Big Data in supply chain context are not addressed in previous studies. Methods to optimise and deploy data generation infrastructure in supply chain network should be explored in future. In recent times, papers have begun to focus on understanding the empirical relationship between the use of advanced analytics and supply chain performance. However, very few studies have addressed the underlying mechanism through which BDA can be utilised to support decision-making and business performance.The positive impact of BDA on business performance is not certain and the influence of several organisational factors remains ambiguous (Oliveira et al. 2012).

Chapter 3 Theoretical Framework

3.1 Chapter introduction

This chapter consists of two parts as shown in Figure 3.1. First, the chapter reviews suitable theories relating to BDA practice and competitive advantage. It covers the theoretical lens used to observe the phenomenon of BDA practice. Then, the rationale for choosing Resource-Based Theory (RBT), Dynamic Capabilities View (DC) and hierarchical of capabilities view as theoretical lenses are provided. In the previous chapter, the research aims related to the BDA practice and value creation is derived from a comprehensive literature review. Broadly, the objectives are two fold 1. To investigate the role of BDA maturity on improving operational and innovation performance, 2. To investigate the role of Absorptive Capacity (ACAP), Data and Information Quality (DIQ) and Supply Chain Analytics (SCA) capability on the previously stated relationship. Consequently, to achieve these aims, in this chapter a conceptual model is developed based on the chosen theoretical lenses to empirically investigate the relationship between BDA Maturity, ACAP, DIQ and SCA capability, and operational and innovation performance. Further, several hypotheses are proposed based on the conceptual model which is also discussed in this chapter.

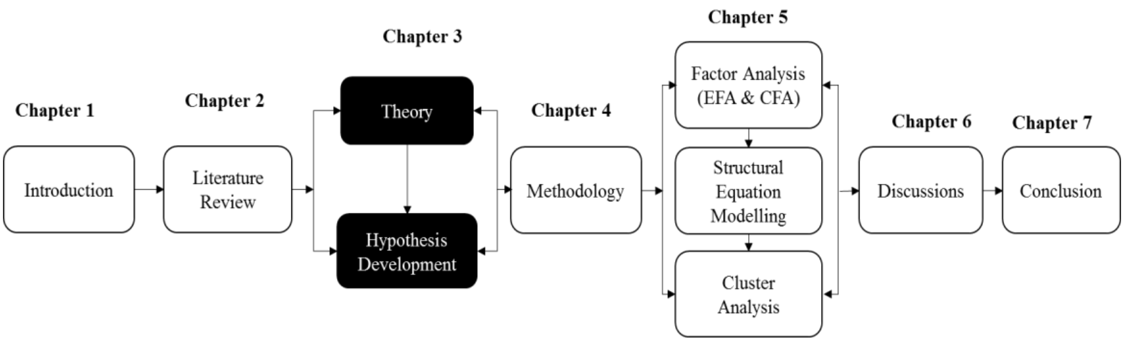


Figure 3.1 Position of the chapter in this thesis

3.2. Theoretical foundation

Kerlinger (1986, p.9) defined theory as ‘‘a set of interrelated constructs, concepts, definitions, and propositions that present a systematic view of phenomena by specifying relationships among variables, with the purpose of explaining and predicting the phenomena’’. So, the four components that the theory is made up of are: 1. theoretical definitions of variables such as manufacturing lead time, 2. domain of a theory- refers to application circumstances, 3. a set of relationships of variables, and 4. predictions (Wacker 1998). Accordingly, this chapter attempts to provide theoretical justifications and proposes a relationship between several variables that would systematically delineate the phenomenon of BDA practice and business value creation. In this research, theory building occurs in a deductive manner by "incremental revision or extension (or occasionally, rejection) of the original theory" (Gioia and Pitre 1990, p.590).

The literature review chapter showed a need for examining the ability of BDA to create business value. Research on the business value of Information Systems (IS) is comprehensive in the academic literature (Mithas et al. 2011; Dale Stoel and Muhanna 2009). BDA is a novel innovative technology related to the information systems discipline but emerging as a separate discipline of its own. In business organisations, technology know-how is often considered as a source of competitive advantage (Powell and Dent-Micallef 1997). It can be argued that organisations are motivated to adopt innovative technologies such as BDA, cloud computing and Internet of Things (IoT) to take advantage of it before their competitors can imitate. It is widely accepted that only by means of developing resources that are Valuable, Rare, Inimitable, and Organisation (VRIO) can an organisation gain a competitive advantage (Barney and Clark 2007; Rothaermel 2017; Brennan 2011). The concept of VRIO resources originates from the Resource-Based View (RBV) and is extensively used in IS literature to explain the role of VRIO resources in enhancing firm performance (Kraaijenbrink et al. 2010; Bowman and Ambrosini 2003; Barney 2001; Kor and Mahoney 2004). However, there is an ongoing debate between the effectiveness of static vs dynamic resources or capabilities, and the theory of Dynamic Capabilities View (DC) as a potential framework to explain the technology facilitated value creation phenomenon in the organisational context. Moreover, from the literature, it is identified that few researchers have conceptualised BDA using both RBV (Gunasekaran et al. 2017; Kwon et al. 2014; Gupta and George 2016) and DC (Wamba et al. 2017; Chen et al. 2015) theory. In this chapter, the rationale for utilising RBV and DC theory are reviewed in the following sections to demonstrate the suitability of these theories to the context of this research.

3.2.1 Resources, capabilities and competencies: An Overview

In the strategy management literature, it is argued that the source of competitive advantage can be present in organisations resources, capabilities, and competencies. There is a lack of clarity in defining these terms (Peppard and Ward 2004), and this section attempts to induce clarity over it. Barney (1991b) perceived resources as inclusive of “all assets, competencies, organisational processes, firm attributes, information, and knowledge that enables a firm to conceive of and implement strategies that improve its efficiency and effectiveness”. Resources are “those assets that deliver value added in the organisation”(Lynch 2015, p.112). Wade and Hulland (2004, p.109) defined resources “as assets and capabilities that are available and useful in detecting and responding to market opportunities or threat”. Hill et al. (2015) referred ‘resources’ as a competitive asset of an organisation that can be classified into tangible and intangible assets. Barney (1991b) recognised physical capital, human capital and organisation capital as different types of resources. The resources that have physical attributes such as manufacturing plants, equipment, and building are called tangible resources (Rothaermel 2017). Whereas, resources such as culture, knowledge, and intellectual properties that are not physical entities, are considered as intangible resources. Rothaermel (2017) argued that in comparison to tangible resources, intangible resources are more likely to produce competitive advantage.

On the other hand, researchers have expressed confounding views when it comes to the narration of capabilities and competencies. For instance, Amit and Schoemaker (1993, p.35) referred to capabilities as “a firm's capacity to deploy resources, usually in combination, using organisational processes, to effect a desired end”. Gamble et al. (2014, p.70) stated that ‘capability’ “is the capacity of a firm to competently perform some internal activity. A capability may also be referred to as a competence”. Hill et al. (2015) have referred to capabilities as “a company’s resource-coordinating skills and productive use”. Similarly, Lynch (2015, p.112) stated that “capabilities of an organisation are those management skills, routines and leadership that deploy, share and generate value from the resources of the organisation.” Barney and Hesterly (2012, p.66) have stated that “capabilities are a subset of a firm's resources and are defined as the tangible and intangible assets that enable a firm to take full advantage of the other resources it controls”. From these logics, consensus on what is resources and capabilities can be arrived at. Capabilities are those that are developed because of effective utilisation of resources and are indeed a subset of resources. Further, there are different kinds of capabilities described in strategic management literature such as organisational capabilities, operational capabilities, and dynamic capabilities. *Organisational capabilities* are defined based on a process perspective as “information-based tangible or intangible processes that are firm-specific and are developed over time through complex interactions among the firm’s resources” (Amit and Schoemaker 1993, p.35). Winter (2003) defined *organisational capabilities* from the perspective of routines as “a high-level routine or collection of routines that, together with its implementing input flows, confers upon an organisation’s management a set of decision options for producing significant outputs of a particular type (p.991)”. In the context of IT research, researchers have conceptualised IT capabilities as an organisational capability, composed of “complex bundles of IT-related resources, skills and knowledge, exercised through business processes, that enable firms to coordinate activities and make use of the IT assets to provide desired results” (Dale Stoel and Muhanna 2009, p.182). Moreover, Wheelen et al. (2018, p.166) argued that “when these capabilities are constantly being changed and reconfigured to make them more adaptive to an uncertain environment, they are called dynamic capabilities”. A detailed discussion of dynamic capabilities is provided in section 3.1.1.

Apart from these, researchers have also discussed another type of capability specific to operations system, i.e. operational capabilities. Operational capabilities are defined as “firm-specific sets of skills, processes, and routines, developed within the operations management system, that are regularly used in solving its problems through configuring its operational resources” (Wu et al. 2010, p.726). Operational capability provides “the means by which a firm functions or operates to make a living in the present” (Brusset 2016) or “how we earn a living now" (Cepeda and Vera 2007, p.427). In contrast to the above definitions, Liu et al. (2013) referred to operational capabilities as “firm's ability to perform operational activities together with channel partners in order to adapt or respond to marketplace changes in a rapid manner”. While the definitions of Brusset (2016), Cepeda and Vera (2007), and Wu et al. (2010) are reflecting internal operations processes and routines of an organisation, Liu et al. (2013) views operational capability at supply chain level, for instance, the authors have contended that supply chain agility is an operational capability. Moreover, scholars refer to operational capabilities as high-level routines (Cepeda and Vera 2007; Brusset 2016; Teece et al. 1997) that can be utilised to respond to market changes (Pavlou and El Sawy 2006; Barreto 2009). It is further referred to as a firm’s ability to coordinate and perform tasks such as distribution logistics and operations planning (Brusset 2016).

Whereas, in terms of competency, Hill et al. (2015) argued that for firms to generate distinctive competencies it is imperative to have both resource and capabilities. Wheelen et al. (2018, p.166) stated that “a competency is a cross-functional integration and coordination of capabilities. The authors further mentioned that the organisation may not require valuable, firm-specific resources if they possess unique capabilities that are not owned by their competitors. Similarly, as shown in Figure 3.2, capabilities are the consequents of integrating resources and competencies (Boddy 2014). Thus, the preceding discussions highlight the importance of focusing on capabilities possessed by organisations compared to resources and competencies.

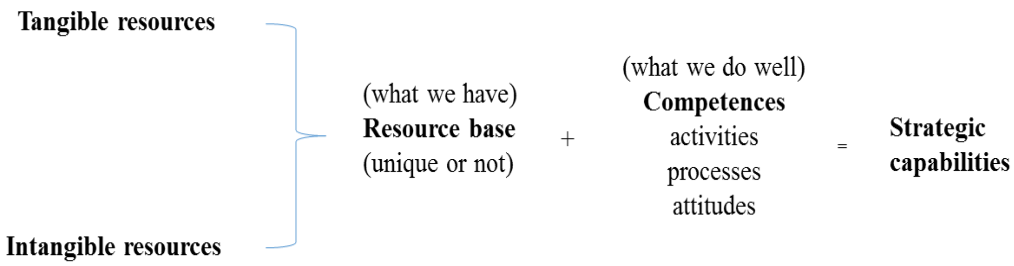


Figure 3.2 Resources, competencies and capabilities

Source: adapted from (Boddy 2014)

3.2.2 RBV and Dynamic capabilities

Understanding the sources of sustained competitive advantage is the central focus of the strategic management discipline (Mata et al. 1995; Porter 1985). The idea of the Resource-Based View (RBV) originates from the works of Edith Penrose who published a book named ‘The Theory of the Growth of the Firm’(Kor and Mahoney 2004; Kor and Mahoney 2000; Penrose 1959). On the objective of understanding the factors that influence firm growth, Penrose (1959) as cited in (Barney and Clark 2007, p.12) identified that firms’ growth is limited based on “the bundle of productive resources controlled by a firm and the administrative framework used to coordinate the use of these resources”. The seminal works of Barney (1986, 1991a) and Peteraf (1993) have further elaborated on the understanding of RBV theory.

Consequently, through adopting RBV, a larger number of IS researchers have identified that IT resources act as a key source of competitive advantage (Bharadwaj 2000). Theorists that adopt RBV contend that physical assets of an organisation can bring competitive advantage only if they “out-perform” comparable assets of their competitors (Barney 1991a; Bharadwaj 2000). It is often argued that physical IT systems can be replicated without difficulty by competitors, and are therefore unlikely to enhance competitive advantage. Similarly, it can be argued that Big Data technology resources can only provide a temporary competitive advantage, as it perishes once the competitors start to duplicate the resources.

As discussed in Mata et al. (1995), the two underlying assumptions of RBV are: (a) resource heterogeneity, (b) resource immobility. Resource heterogeneity denotes a scenario in which all firms within an industry possess different resources and capabilities compared to their competitors. It is illogical to assume that all firms in an industry can possess identical resources (Barney 1991b). If all firms in an industry have homogenous resources, no firm would gain a sustainable competitive advantage. However, this notion leads to the concern of ‘first mover advantages’, i.e. only the first firm in an industry to adopt a resource or implement a strategy would gain a competitive advantage. Resource immobility indicates that “resources tend to be “sticky” and don’t move easily from firm to firm” (Rothaermel 2017, p.112). It is argued that if competing firms do not possess a capability and find it difficult to adopt it due to cost disadvantage (i.e. resource immobility), then the firm that possesses the capability can gain a sustainable competitive advantage (Mata et al. 1995). Barney and Clark (2007) stated some of the reasons for resource immobility: (a) the organisation’s ability to generate a capability in a profitable manner depends on the firm’s historical conditions, (b) the creation of a capability could be path-dependent - the antecedents of capability may be unknown and involve a difficult learning process, making it hard to replicate, (c) the complexity of creating a capability is high due to intricate social structures (e.g. a firm’s culture), (d) the procedures, actions or activities needed to develop a capability may be unknown to firms. Indeed, if the cost of creating a capability is superior to the value it would bring, rationally firms would lack self-interest in replicating a capability, and it provides a long-lasting competitive advantage to firms who possess such a capability (Barney and Hesterly 2012a).

These two basic assumptions of RBV theory led to the development of the VRIO framework as shown in Figure 3.3. The VRIO framework discussed earlier is used as a tool for the analysis of resource and capabilities. Accordingly, in order for a resource or capability to be considered as a source of competitive advantage, the resource should be unique, i.e. valuable, rare, costly to imitate, and organised effectively within an organisation (Rothaermel 2017; Oh et al. 2014). A valuable capability would allow an organisation to exploit external opportunities and neutralise threats (Rivera and Shanks 2015; Rothaermel 2017), create economic value (Barney and Clark 2007), and then develop and implement strategies to enhance efficiency and effectiveness. A rare capability is the one that is not possessed by any competitors (at least relatively not at the same level) as the focal firm (Wheelen et al. 2018). An imperfect imitable capability is something that the competing firms are unable to imitate due to historical conditions, causal ambiguity, and social complexity (Barney and Clark 2007), and costly to imitate (Rothaermel 2017). The fourth condition, 'Organisation', refers to the internal structure of firms. It stresses the need for organising the valuable, rare, and non-imitable capability via management support, internal structure and coordinating systems for exploiting the capability to achieve a sustainable competitive advantage. Based on these RBV and VRIO logic, resources/capabilities that are not valuable, not rare, easily imitable, and not organised would lead to a competitive disadvantage; resources/capabilities that are valuable but not rare, easily imitable, and not organised would lead to competitive parity; resources/capabilities that are valuable and rare but easily imitable and not organised can serve as a source of a temporary competitive advantage; and resources/capabilities that satisfied all the four parameters (i.e. valuable, rare, costly to imitate, and organised to extract value) can bring a sustainable competitive advantage (Rothaermel 2017; Barney and Clark 2007; Dale Stoel and Muhanna 2009). This reinforcement of RBV has led organisations to intensely focus on developing 'unique' resources. However, over time, this static view of considering a firm's resources base as the antecedent of competitive advantage (resource possession) is expanded to focus on resource utilisation (Fawcett et al. 2011; Priem and Butler 2001). Thus, RBV argues that equal importance should be given to the possession of a unique resource base as well as to developing and reconfiguring these resources/capabilities to make the best use of its competitive potential (Eisenhardt and Martin 2000). To summarise, resources that do well on the VRIO assessment can deliver a sustainable competitive advantage, either by product advantage-delivering superior value to customers or by process advantage-decreasing the unit costs (Bowman and Ambrosini 2003). As such, it generates economic rents and improves the profitability of firms.

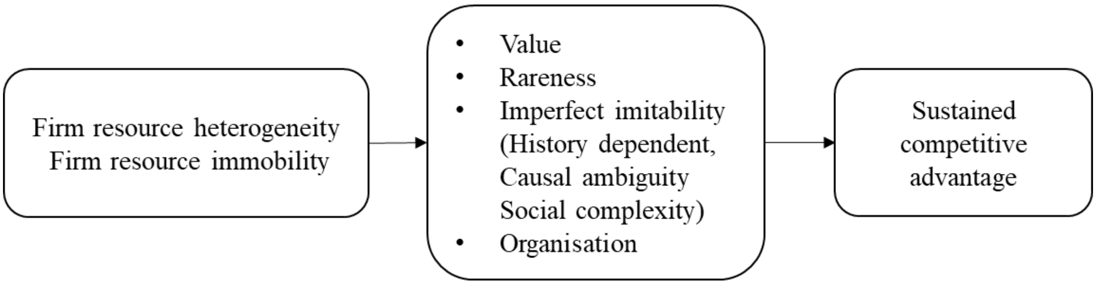


Figure 3.3 Relationship between assumptions of RBV and VRIO factors

Source: (Barney and Clark 2007)

However, RBV is not left without criticism by academic scholars. For instance, (Kraaijenbrink et al. 2010) reviewed the theory and discussed several worthwhile criticisms, as shown in Table 3.1. RBV is critiqued as a static theory (Priem and Butler 2001), as it focuses on recognising resources at one point in time (Bowman and Ambrosini 2003), and "focuses on the internal organization of firm" (Eisenhardt and Martin 2000, p.1105). RBV assumes that even static resources would lead to a sustained competitive advantage. This claim is criticised and disputed considering growing uncertain and dynamic business environment (Eisenhardt and Martin, 2000; Cosic et al. 2015; Olszak, 2016). Moreover, it is argued that attaining competitive advantage is contingent upon reconfiguration of resources.

Table 3.1 Important critiques to the Resource-Based View (RBV) and assessment

Critique	Assessment /reflection
The RBV is claimed to lack managerial implications (Priem and Butler 2001). RBV suggests managers develop VRIO Resources but no implications are offered on how to develop it.	Provision of managerial implications directly is not a prerequisite of any theories. Moreover, RBV has an evident impact on the managerial practice. Hence this critique can be trivial.
The RBV indicates infinite regress(Collis 1994), suggesting firms should attempt to acquire second-order capability (as it is a potential source of competitive advantage) than the first-order capability. This view is critiqued as it may lead firms to infinite search for higher-order capabilities.	But, this critique may be relevant to abstract mathematical theories. RBV is an applied theory, the different levels of capabilities are qualitatively different. Moreover, it is beneficial to focus on the interactions between different capability level.

The sustainable competitive advantage is not achievable (Fiol 1991; Fiol 2001)	"By including dynamic capabilities, the RBV is not purely static, though it only explains ex-post, not ex-ante, sources of sustainable competitive advantage. Although no competitive advantages can last forever, a focus on sustainable competitive advantage remains useful" (Kraaijenbrink et al. 2010).
The RBV is claimed to be a new theory of the firm, but this view is critiqued as RBV's agenda is different from theories of the firm such as Transaction Cost Economics (TCE)	The RBV is adequate in clarifying why firms exist. But, RBV should be developed as a theory of sustainable competitive advantage rather than trying to be a theory of the firm.
The definition of resource is impractical (Conner 1991; Conner and Prahalad 1996; Wade and Hulland 2004). The definition of 'resource' by RBV theorists seems to be over-inclusive. As a result, there is a lack of acknowledgement of differences between resource and capability. s	There is no acknowledgement of ways through which different resources contribute to sustainable competitive advantage.

Source: adapted from (Kraaijenbrink et al. 2010)

In response to these criticisms on RBV, Dynamic Capabilities View (DCV) perspective has emerged in strategic management discipline and commonly applied in IS and supply chain research. The rationale for extending RBV into DCV is that RBV does not sufficiently explain "how and why certain firms have a competitive advantage in situations of rapid and unpredictable change" (Eisenhardt and Martin 2000, p.1106). In the landscape of a rapidly changing environment, the use of dynamic capabilities is considered as the main source of competitive advantage. According to Teece et al. (1997), dynamic capabilities is "the firm's ability to integrate, build, and reconfigure internal and external competencies to address rapidly changing environments"(p.516). Some definition of dynamic capability is provided in Section 2.6.5.1 in the context of Absorptive Capacity. A more generic definition of dynamic capabilities widely used in literature is given below.

"*The firm's processes that use resources—specifically the processes to integrate, reconfigure, gain and release resources—to match and even create market change. Dynamic capabilities thus are the organizational and strategic routines by* *which firms achieve new resource configurations as markets emerge, collide, split, evolve, and die*" (Eisenhardt and Martin 2000, p.1107).

In contrast to RBV, it is argued that Dynamic capabilities view (DCV) "addresses the process of future resource creation" (Bowman and Ambrosini 2003, p.292), bringing in an external perspective (Eisenhardt and Martin 2000). In a dynamic and complex environment like a manufacturing supply chain, assets (e.g. BDA technology assets) alone are not sufficient to bring competitive advantage, but organisational capabilities to utilise these BDA resources are important to gain competence (Opresnik and Taisch, 2015). Dynamic Capabilities (DC) are essential to constantly renew, recreate and reconfigure resources and capabilities to address the changing environmental needs and to attain lasting firm performance (Teece et al. 1997; Opresnik and Taisch, 2015).

Moreover, scholars have recommended that organisational capabilities can be conceptualised as a hierarchy (Kraaijenbrink et al. 2010; Grewal and Slotegraaf 2007; Grant 1996a; Sirmon et al. 2007). It is argued that higher-order capabilities are difficult to mimic by competitors, and are developed by a sequence of lower-order capabilities. Grewal and Slotegraaf (2007, p.155) have elaborated the perspective of the hierarchy of capabilities: at the foundation level there are "firm's specialized knowledge and resources, which are combined to generate lower-order capabilities; these, in turn, are combined to generate higher-order capabilities". Grant (1996a) has illustrated the hierarchy of capabilities with an example: a manufacturing firm generates engineering capability utilising its specialised knowledge and resources, which in turn integrated with lower-order capabilities produces operations capability, this, in turn, is integrated with functional capabilities such as R&D capabilities to produce a new product development capability. Accordingly, in information systems research, academics increasingly regard information systems capabilities as a lower-order capability that allow the development of higher-order capabilities such as supply chain agility (Sambamurthy et al. 2003; Liu et al. 2013), knowledge management (Tanriverdi 2005), dynamic and operational capabilities (Pavlou and El Sawy 2006), and supply chain process integration capability (Rai et al. 2006). Sambamurthy et al. (2003) and Rai et al. (2006) have argued that higher-order capabilities are the significant source of sustainable competitive advantage. Following these logics, this study takes the view of hierarchy of capabilities, and proposes that BDA capabilities maturity is a lower-order capability that enabled the development of higher-order capabilities such as ACAP, DIQ and SCA capabilities, that, in turn, will directly affect operational and innovation performance as shown in the conceptual model (Figure 3.6).

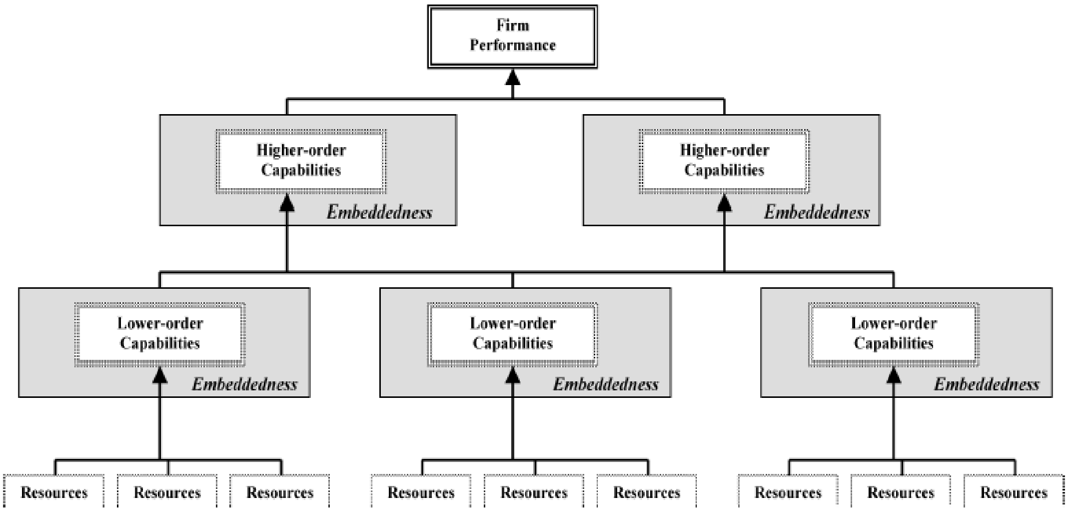


Figure 3.4 illustration of organizational capabilities: A multi hierarchy perspective

Source: (Grewal and Slotegraaf 2007)

Hence, based on the theoretical arguments, the phenomenon of Big Data Analytics (BDA) capabilities and its impact on operational and innovation performance is viewed through the lens of Resource-Based View (RBV), DCV, and hierarchy of capabilities. Increasingly, nowadays, organisations consider the BDA capabilities as a strategic resource to attain competitive advantage, and this research draws upon RBV to explain the extent to which BDA resources are VRIO and contribute to business value. From Figure 3.5, it is evident that RBV is widely accepted in the domain of BDA research. RBV is highly influential in the strategic management discipline and used to investigate the competitive advantage derived from resource competence (Sangari and Razmi, 2015).

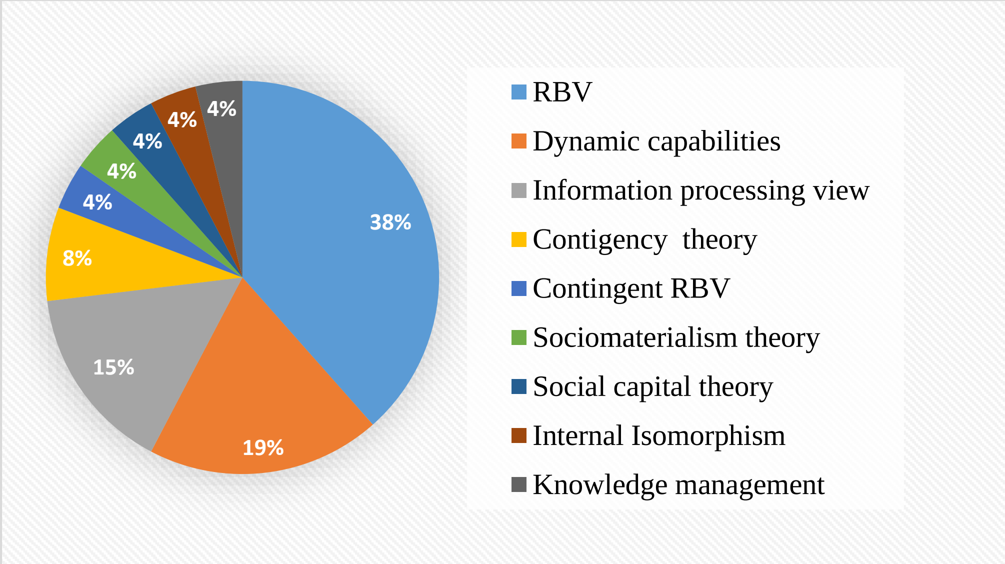


Figure 3.5 Prominent theories used in BDA literature

Moreover, based on the views of RBV and DCV, BDA capabilities maturity is perceived as a lower-order organisational capability; absorptive capacity is regarded as higher-order dynamic capabilities; Data and information quality is perceived as a higher-order organisational capability; supply chain analytics capability is considered as a higher-order operational capability. This is because organisations that possess a bundle of Big Data resources such as advanced databases, analytical tools, and skilled employees will reconfigure them in a unique way to create BDA capabilities. The opportunities to develop these firm-specific static BDA capabilities are common to all firms in the UK manufacturing industry provided they have no financial constraints. So, from the DC view, it can be argued that the use of BDA can develop organisations' information processing capabilities, facilitate resource reconfiguration, reduce uncertainties, and predict future resource requirements (Chen et al. 2015a). Consequently, a conceptual model is developed through the lenses of RBV and DC and is discussed in the following section.

3.2.3. Conceptual model development

Figure 3.6 shows the conceptual model developed in this study. The model illustrates the relationship between the variables BDA capability Maturity, Absorptive capacity(ACAP), Data and Information Quality (DIQ), Supply Chain Analytics Capabilities (SCA), and operational and innovation performance. Further, based on the literature review, BDA capabilities maturity is perceived as a multidimensional second-order construct consisting of seven first-order constructs such a Data

Generation Capabilities (DG), Data Integration and Management Capabilities (DIM), advanced analytics capabilities (AA), Digital Analytics Capabilities (DG), Data Visualisation Capabilities (DV), Data-Driven Culture (DC), and Big Data Skills (BDS) (Arunachalam et al. 2018; Gupta and George 2016). Similarly, ACAP is also conceptualised as a multidimensional construct composed of four dimensions: acquisition, assimilation, transformation, and application (Ali and Park 2016; Zahra and George 2002; Hefu Liu et al. 2013b). DIQ and SCA are conceptualised as a unidimensional first-order construct. Moreover, as discussed in the literature review chapter, scholars have considered operational performance as a multidimensional construct (Gunasekaran and Kobu 2007). In this research, four different operational performance dimensions which are frequently used in literature as indicators are included in the conceptual model such as product quality, cost-based performance, flexibility and time-based performance. These dimensions are identified from the literature review of operational performance discussed in section 2.7. Besides, innovation performance is also included in the model as a unidimensional construct.

In the conceptual model, BDA maturity is the exogenous variables influencing endogenous variables such as ACAP, DIQ, SCA, operational and innovation performance. Moreover, ACAP, DIQ, and SCA are also the mediating variables in the conceptual model. In addition, firm attributes such as a number of employees and annual turnover are included in the model as control variables. This is because, firms which are highly dynamic in nature (e.g. High-tech manufacturing) would face volatility in terms of revenue generation and customer retention than less dynamic companies (Wu and Chuang, 2010; Narasimhan and Kim, 2002). Also, larger firms tend to adopt BDA technology more than smaller firms due to their economic competence to possess the right resources and the right skills. So, the relationship between the variables of the conceptual model may vary based on firm attributes and hence control variables are also incorporated in the model. Moreover, many previous studies such as Gupta and George (2016), Wagner (2006) and Yu et al. (2013) have considered these firm attributes as control variables in their model. Thus, the research model signifies the key BDA capabilities that organisations should possess so as to enhance operational and innovation performance, and the conceptualisation will eventually provide guidance for companies seeking to develop BDA capabilities.

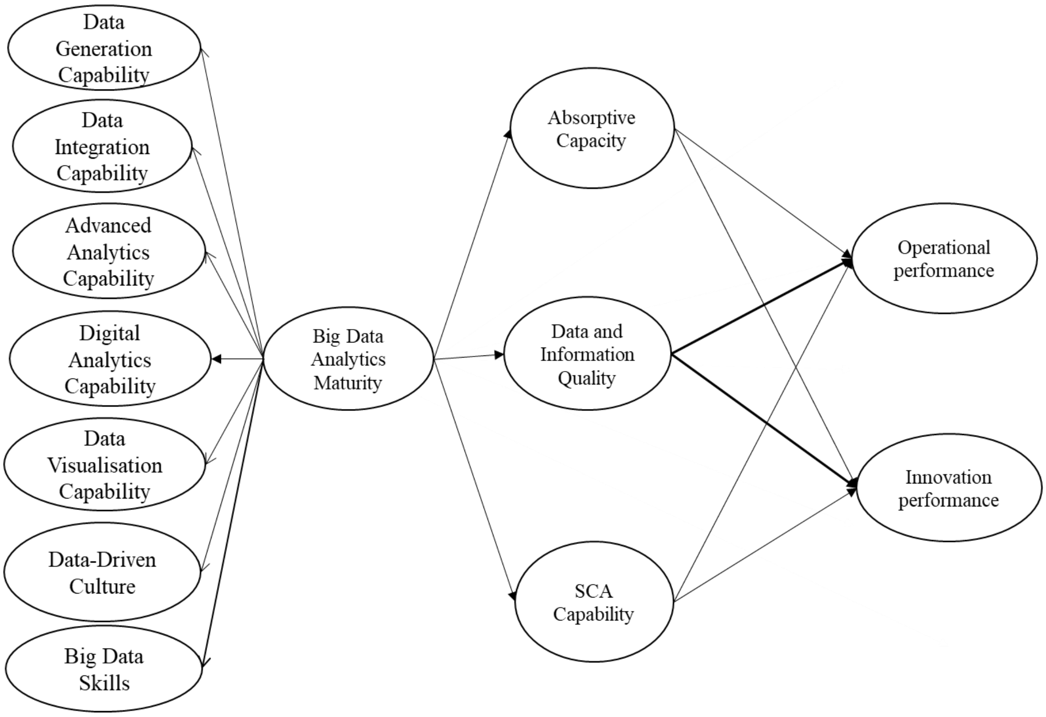


Figure 3.6 The conceptual model of this research

3.3 Development of research hypotheses

3.3.1 Direct effect of BDA capabilities and operational and innovation performance

The performance impact of BDA capabilities has gained significant interest among practitioners and academic researchers. Foremost, to operationalise BDA capabilities Maturity construct, it is hypothesised that the seven first-constructs DGC, DIMC, AAC, DAC, DVC, BDS, and DDC are the reflection of BDAM and the validity and the reliability of the second-order construct has to be tested. However, this section mainly focuses on discussing empirical evidence that supports the relationship between BDA maturity and operational and innovation performance.

The literature on BDA revealed various benefits of practising BDA at manufacturing firms (intra-organisational level). Few scholars have demonstrated the influence of utilising Big Data Analytics at the organisational level on operational performance. For instance, Chae, Yang, Olson, et al. (2014), empirically investigated the impact of analytics on improving manufacturing performance. The authors found out that advanced analytics capabilities of organisations measured at organisational level have a significant positive effect on operational performance measured on variables such as order fulfilment, delivery as promised, delivery flexibility, flexibility to change output volume, and flexibility to change product mix. Similarly, Cao et al.

(2015) measured the impact of various business analytics techniques including social media and web analytics on the effectiveness of operational decision-making in manufacturing firms and found out that there is a significant positive relationship between them. Schmidt et al. (2014) have demonstrated how a cloud-based analytics platform can be used to optimize price, time and quantity to save cost and reduce risk. Tao and Storey (2015) have used user-generated content to create a recommendation system that is capable of reducing costs related to products and services. Evidence on the usage of BDA for reducing contamination risks and assurance of food quality is also available in the literature (Ting et al. 2014). Moreover, Big Data offers several benefits such as the integration of qualitative and quantitative data, improved observation of events, better efficiency and effectiveness of systems, and evidence-based decision-making (Tien 2012).

Furthermore, the installation of RFID tags and readers on logistic objects can convert them into 'passive smart logistics object' and 'active smart logistics object' (Zhong, Lan, et al. 2016). RFID enables the manufacturing shop floor to generate a huge amount of data. When these physical objects start communicating via wireless communication, an enormous volume of data gets generated in real-time making it difficult to handle. For instance, a manufacturing plant that has deployed 1,000 RFID readers and 10,000 tags could potentially generate terabytes of data from a single day of operation (Zhong et al. 2015; Zhong, Lan, et al. 2016). It can be argued that these data contain valuable information and have the potential to improve operational decision-making. Sonka (2014) have stated several instances where BDA can be used to improve manufacturing efficiency. A missile manufacturing plant (Raytheon Co) exploits BDA to monitor deviation and defects in the screwing process, which instantly sends an error message if a malfunction occurs. Similarly, Harley-Davidson constantly measures the physical status of the machine that is in production, especially its temperature, humidity, etc. The system is designed in such a way if these variables deviate from the standard the machine will automatically adjust itself. Moreover, Harley-Davidson also uses BDA to reconfigure operational strategy to increase the efficiency of their assembly line. Toyota enhanced its lean manufacturing techniques based on the data analysis which impacted on transforming the relationship with its suppliers (Sonka 2014). Markham et al. (2015) elaborated various business applications of text analytics capabilities and its benefits such as to identify customer needs, competitors' moves, and also to facilitate new product development decision. Moreover, text analytics capability is widely used to track customer complaints and satisfaction level using data available in a company's complaint logs, social media, and blogs.

In addition, scholars have discussed the potential benefits of analytics tools and techniques to optimise and reduce the cost of inventory (Chuang et al. 2014; Chiang et al. 2011a). For instance, Chiang et al. (2011b) confirmed the benefits of utilising data from Warehouse Management Systems (WMS) to optimise and solve the storage location assignment problem. Chuang et al. (2014) have utilised an association rule mining technique to analyse sales order data to identify product co-occurrences, which allowed them to improve flexibility and customisation of product volumes in real-time. Zheng et al. (2014) developed an integrated data analytics platform to automate and optimise various processes in plasma display panel manufacturing. BDA is also found to be instrumental in the maintenance of manufacturing machinery. Traditional maintenance procedures are mostly based on the aggregated historical data, which causes machines to be over-used and under maintained and results in the production of low-quality products and also affects the production yield (Lee et al. 2013; Babiceanu and Seker 2016). BDA can enable real-time sensing and identification of machine faults and enhances control over quality and production yield. Similarly, Zhang et al. (2017) stated that BDA plays a major role in manufacturing enterprises for improving product innovation and proactive maintenance of manufacturing machinery. Hahn and Packowski (2015) estimated that collection and analysis of data generated from sensor-based technology could substantially reduce the operational cost by 10 to 25%. Metan et al. (2010) used data from a manufacturing database and analysed it with techniques such as decision trees and statistical process control to propose an effective scheduling system that dispatches materials in real-time and enhances operational efficiency.

Further, Dutta and Bose (2015) have illustrated a case study on the implementation of BDA in Ramco Cements Limited (RCL) - a cement manufacturing company. After the implementation of BDA, RCL has achieved several benefits such as, increased market share, sales growth, optimisation of truck unloading time, accuracy of packaging, reduced costs, improved customer satisfaction through reduction in delivery time, better on-the-go sales strategies, improved market penetration strategy, improved access to critical operations information, and overall BDA has significantly increased information processing capabilities of RCL. Manyika et al. (2011) argued that BDA can allow manufacturers to cut down the time of product development by 20 to 50 % and eradicate defects proactively before production through simulation and testing. In addition, Dutta and Bose (2015) have noted that successful BDA adoption and usage depends on managing cultural change, training and improving employees commitment to use these new tools, effective collaboration across different stakeholders of the organisation, and top management support. This implies that development of big data skills and a data-driven culture also play a major role in improving firm performance.

Hence, in this research, the following hypotheses are proposed related to the relationship between BDA maturity and operational performance.

H1: Big data analytics maturity has a significant positive effect on operational performance dimensions.

H1a: Big data analytics maturity has a significant positive effect on product quality performance.

H1b: Big data analytics maturity has a significant positive effect on cost performance.

H1c: Big data analytics maturity has a significant positive effect on flexibility performance.

H1d: Big data analytics maturity has a significant positive effect on time-based performance.

In addition, since investigating the influence of the first-order BDA capabilities on operational performance is also a motive of this study, the following hypotheses are proposed:

H1e: Data Generation capability has a significant positive effect on overall operational performance.

H1f: Data Integration capability has a significant positive effect on overall operational performance.

H1g: Advanced Analytics capability has a significant positive effect on overall operational performance.

H1h: Digital Analytics capability has a significant positive effect on overall operational performance.

H1i: Data Visualisation capability has a significant positive effect on overall operational performance.

H1j: Big Data Skills has a significant positive effect on overall operational performance.

H1k: Data-Driven Culture has a significant positive effect on overall operational performance.

Besides the positive effect of BDA maturity on operational performance, scholars have also reinforced the effectiveness of BDA to influence innovation performance. Big data is considered as the next frontier of innovation (Marr 2016; Manyika et al. 2011). Erevelles et al. (2015) argued that data generated and analysed from social media, mobiles and electronic point-of-sale have great potential to reveal insights about consumer behaviour, product preferences, and expectation. These types of data and analytics provide insights appropriate for innovating new products and services. Similarly, Xu et al. (2015) argued that BDA provides a knowledge-discovery mechanism for mining text data that is unstructured and messy, and firms taking advantage of it can customise new products and services. Tuarob and Tucker (2015) have discussed the case of a smartphone manufacturer utilising BDA to understand the desired features on a smartphone, allowing the manufacturing to innovate new product features and increase customer satisfaction. Wamba et al. (2015) argued that the data, the analytics and the presentation of the results would allow firms to create value in terms of developing new products/services. Tan et al. (2015) demonstrated that the BDA framework based on the deduction graph method can improve innovation capabilities.

Marr (2016) explained the impact of BDA on innovation using real cases. Rolls-Royce, a manufacturer of high-tech aircraft engines, deployed BDA resources in three areas of their business: design, manufacture and after-sales services. In the design phase of engines, the company generates tens of terabytes of data during the simulation of one of their jet engines. Paul Stein, the Chief Scientific Officer (CSO) of Rolls-Royce, revealed that the visualisation of Big Data generated during the design phase, helped them to visualise the behaviour of jet engines and decide whether the product is good or bad. Eventually, it enhanced the company's ability to manufacture new products with improved quality. Moreover, Rolls-Royce's engines are fitted with hundreds of sensors that collect real-time data about any changes in the operation and data engineers can analyse it and offer better services to their customers. In addition, BDA facilitated Rolls-Royce to develop a new business model based on data-driven insights. As a general remark, Paul Stein has stated that BDA-enabled “innovation in service delivery was a game- changer”. It is seemingly likely that improving the quality of decisions will have a significant effect on process management (Ipcc 2013), as well as a better decision over new product development and innovation (Tan et al. 2015)

Hence, in this research, the following hypotheses are proposed related to the relationship between BDA maturity and innovation performance.

H2: Big data analytics maturity has a significant positive effect on Innovation performance.

In addition, since investigating the influence of the first-order BDA capabilities on innovation performance is also a motive of this study, the following hypotheses are proposed:

H2a: Data Generation capability has a significant positive effect on Innovation performance.

H2b: Data Integration capability has a significant positive effect on Innovation performance.

H2c: Advanced Analytics capability has a significant positive effect on Innovation performance.

H2d: Digital Analytics capability has a significant positive effect on Innovation performance.

H2e: Data Visualisation capability has a significant positive effect on Innovation performance.

H2f: Big Data Skills has a significant positive effect on Innovation performance.

H2g: Data-Driven Culture has a significant positive effect on Innovation performance.

3.3.2 Direct effect of BDA capabilities on ACAP, DIQ, and SCA capabilities

Based on the perspective of the hierarchy of capabilities, scholars have suggested that lower-order organisational capabilities such as BDA capabilities and IT capabilities can facilitate firms to build higher-order capabilities (Liu et al. 2011). Consequently, in this research, hypotheses are proposed illustrating the influence of BDA capabilities maturity on absorptive capacity, data and information quality, and supply chain analytics capabilities. As discussed in section 3.1.1, the lower-order capabilities are crucial for developing higher-order abilities, that, in turn, improve firm performance. It can be argued that BDA maturity is a lower-order organisational capability that can enhance a firm's ability to develop ACAP (dynamic capability), DIQ (organisational capability), and SCA capability (operational capability), which are higher-order in nature. Moreover, scholars have empirically identified the interaction between lower-order and higher-order capabilities. For instance, Takeishi (2001) have found out that automakers internal capabilities of coordinating has enhanced inter-firm coordination capabilities, which helped the organisations to improve performance. Also, Pfohl and Buse (2000) argued that intra-organisational logistics capabilities have a significant effect on inter-organisational logistics systems.

First, with respect to the impact of BDA on absorptive capacity, academic scholars have identified that IS capability is an antecedent of absorptive capacity - conceptualised as dynamic capabilities (Malhotra et al. 2005; Wang et al. 2014; Fayard et al. 2012; Noblet et al. 2011; Ettlie and Pavlou 2006). ACAP is the ability of organisations to manage external knowledge and use it to generate commercial value, and the concept is widely applied in IS research (Roberts et al. 2012). Tzokas et al. (2015) have found out that technological capabilities have a significant positive effect on ACAP. Bellamy et al. (2014) have empirically found out that improvement in supply network capabilities (accessibility and interconnectedness) increases the flow of information and knowledge having the positive effect on the R&D intensity of a firm (ACAP). Sáenz et al. (2014) found out that buying firm-specific organisational capabilities have a significant positive effect on improving ACAP. Liu et al. (2013) measured ACAP using four dimensions (such as acquisition, assimilation, transformation, and exploitation) and found out that a firm's IT capabilities help improve absorptive capacity of the firm. Blohm et al. (2013), based on a case study, have found out that the ability to absorb crowdsourced data has a significant potential to improve a firm's ACAP. Further, the electronic integration capability of an organisation is found to have a positive effect on a firm's ACAP (Fayard et al. 2012). Meredith et al. (2012) have found out that the adoption of strategic business intelligence systems a key enabler of dynamic capabilities such as ACAP. Similarly, it can be argued that firms with better BDA capabilities could benefit from the effective acquisition of external knowledge and would improve ACAP. Thus, based on a hierarchy of capabilities and DC view, the following hypothesis is proposed which states that firms with higher levels of BDA capabilities would have enhanced ACAP (dynamic capabilities).

H3: Big data analytics maturity has a significant positive effect on Absorptive capacity

Second, the impact of BDA on Data and Information Quality (DIQ) has received significant interest among the academic scholars (Merino et al. 2016; Song et al. 2017; Ryu et al. 2006; Wang and Strong 1996; Warth et al. 2011; Chae, Yang, Olson, et al. 2014; Popovič et al. 2009; DeGroote and Marx 2013; Haryadi et al. 2016; Nelson et al. 2005). Strong (1997) found out that the IT process design capability has a significant effect on data quality based on a simulation experiment. Popovič et al. (2012) measured Business Intelligence Maturity (BIM) using two reflective constructs (data integration and analytical capabilities) and found that BIM maturity has a significant positive effect on data quality and media content quality. Gorla et al. (2010) empirically found that information systems quality has a significant positive effect on the information quality of an organisation. The Customer Relationship Management (CRM) infrastructure capability is found to have a positive influence on enhancing the quality of customer information available to organisations (Chuang and Lin 2013). Chae, Yang, Olson, et al. (2014) have identified that advanced analytics along with data accuracy has a significant impact on improving the quality of planning in manufacturing organisations. Warth et al. (2011) validated that IT usability has a positive effect on enhancing data and information quality. Further, Kenett and Shmueli (2017) stated that “integrating multiple sources and/or types of data often creates new knowledge regarding the goal at hand, thereby increasing data and information quality” (p.37). Lee et al. (2006, p.4) argued that often traditional data warehouse and ERP systems failed to deliver business values due to the lack of ability to maintain data quality. Lee et al. (2006, p.4) stated this scenario with an instance, a global manufacturing company had tried to combined data and information about their sales across the globe. Although the company generates enormous raw data, it took several months for them to integrate and deliver a usable dataset that can satisfy their business needs. The main cause of the problems is the lack of access to physical databases and lack of a standard procedure to maintain data quality in terms of accuracy, reliability, completeness, and timeliness. However, due to the advent of BDA, it can be argued that the organisation's ability to process raw data using advanced data integration and management tools, can improve data and information quality. Hence, in this research, the ability to maintain data and information quality is treated as a higher-order organisational capability enhanced by organisations' BDA capabilities maturity. The following hypothesis is proposed based on the above arguments and the hierarchy of capabilities view:

H4: Big data analytics maturity has a significant positive effect on data and information quality

Third, the interaction between the usage of BDA at the intra-organisation level and inter-organisation level has received limited attention in the academic literature. However, IS scholars have noticed that the organisations that are fully adept at utilising a technology or practices at the intra-organisational level are more likely to deploy it at the supply chain level. Luzzini et al. (2015) argued that organisational commitment may lead to the development of both intra- and inter-organisational capabilities to realise the greater competitive advantage. Based on the hierarchy of capabilities, scholars have witnessed the impact of intra-organisational IT capabilities on inter-organisational operational capabilities such as supply chain agility (DeGroote and Marx 2013; Hefu Liu et al. 2013b) and information processing capabilities (Oliveira et al. 2012). Fayard et al. (2012) found out that an organisation's internal electronic integration and cost management capabilities have a significant positive relationship with inter-organisational cost management capabilities. Meredith et al. (2012) claimed that business intelligence system should gradually mature to include organisation wide internal systems and then reach out to include external systems. Similarly, it can be argued that a higher level of BDA maturity at intra-organisational level could enhance an organisation's ability to utilise it at an inter-organisational level to extract full value from the investment. Moreover, an organisation's ability to reconfigure and utilise it across various functional areas such as the supply chain, are difficult to imitate by competitors, as they are tacit knowledge that eventually could create a sustainable competitive advantage. Thus, the organisation's increased level of expertise in BDA practice for intra-organisational decision-making processes could be the antecedent of data-enabled supply chain practices. Hence, based on the above arguments and using the hierarchy of capabilities perspective, the following hypothesis is proposed:

H5: Big data analytics maturity has a significant positive effect on supply chain analytics capability

3.3.3 The mediation role of ACAP in the relationship between BDA maturity and operational and innovation performance

In this research, ACAP is conceptualised as a dynamic capability referring to an organisation's ability to acquire, transform and apply external knowledge to improve firm performance. Several scholars have attempted to investigate the consequents of ACAP, especially in the context of technology usage and innovation. Ali and Park (2016) re-conceptualised Absorptive capacity into two-dimensions: 1) Potential Absorptive capacity (PACAP), 2) Realised Absorptive capacity (RACAP) and investigated the sequential influence of both dimensions of ACAP on organisational innovation. The authors found out that the direct effect of absorptive capacity on the product, process and management innovation is significant, but mediated by the innovation culture. Similarly, an empirical study conducted in the context of high-performance manufacturing revealed that ACAP has a positive effect on innovation, delivery, quality, cost and flexibility dimensions of operational performance (Beheregarai Finger et al. 2014). Further, the Beheregarai Finger et al. (2014) study also showed that ACAP positively mediates the relationship between supply chain planning and operational performance. Blohm et al. (2013) found out that ACAP enabled by crowdsourced data has the potential to create business value via innovation. In addition, absorptive capacity is also proficient in improving Inter-organisational costs management practices thereby increasing the cost-based performance of an organisation (Fayard et al. 2012). Jabar et al. (2011) empirically verified the positive influence of ACAP on technology transfer and new product development in a manufacturing context. ACAP is also found to negatively influence B2B e-commerce risk, so organisations with an increased level of the absorptive capacity can significantly reduce risks (Arnold et al. 2011). Based on these arguments, this research suggests that ACAP is a significant source of greater operational and innovation performance. This is because firms with high level of ACAP can effectively manage external knowledge obtained from BDA practice and apply it to identify new business opportunities, customer preference, etc., and subsequently can improve profitability and market share (Liu et al. 2013b). Further, ACAP ensures the reach of external knowledge across diverse functional departments of the firms, facilitating firms to successfully apply the knowledge and improve operational excellence. So, the following hypotheses are proposed:

H6a: Absorptive capacity is positively related to operational performance.

H6b: Absorptive capacity is positively related to innovation performance.

As discussed in section 3.2.2, in comparison to lower-order capabilities, resources provided by higher-order capabilities are difficult to imitate (Grant, 1996). In this study, it is hypothesized that while capabilities related to BDA can directly influence firm performance, a robust model would require ACAP as a mediating factor contemplating an indirect effect. While the effect of ACAP is not investigated empirically in the context of BDA, there are several studies that relate ACAP to supply chains and IT, which supports our argument. The Systematic Literature Review (SLR) of the ACAP literature revealed that ACAP is the consequents of information systems capabilities and plays a mediating role in explaining causal relationships. For instance, in DeGroote and Marx (2013) and Liu et al. (2013b), ACAP is hypothesised as a mediator for explaining the impact of IT capabilities on operational performance. Studies have also found ACAP plays a significant role in explaining the variation in innovation performance (Wang et al. 2014; Liao et al. 2010). Tzokas et al. (2015) investigated the relationship between technical capabilities and customer relationship capabilities on firm performance mediated by ACAP. ACAP is found to positively mediate the relationship between technical capabilities and firm performance. Dobrzykowski et al. (2015) investigated the role of absorptive capacity on the relationship between a firm's responsive strategy and performance and found that ACAP has a significant direct relationship with a firm’s performance as well as mediating the relationship between responsive strategy and performance dimensions. Thus, the following hypotheses are proposed depicting the mediating role of ACAP in the relationship between BDA Maturity and operational and innovation performance:

H7: Absorptive capacity mediates the relationship between BDA maturity and operational performance.

H8: Absorptive capacity mediates the relationship between BDA maturity and innovation performance.

3.3.4 The mediating role of DIQ in the relationship between BDA maturity and operational and innovation performance.

Data and information quality is an important issue, as poor data quality can have a disastrous consequence (Woodall et al. 2013; Redman 1998; Redman 2001; Redman 1996). The availability of quality data could directly affect the process management. It can inform employees about changes in the processes immediately so corrective action can be taken in a timely manner (Kaynak 2003). Malhotra et al. (2005) argued that the incompleteness of data may negatively influence the decision-making effectiveness.

Big data analytics capabilities, especially data integration and management capability, can improve data quality by acquiring and integrating data from various sources to provide a single point of truth (Arunachalam et al. 2017). BDA can improve data quality by utilising its raw data processing capabilities. This is because raw data could inherently contain irregularities due to flawed system design and data input errors. The absence of BDA can create a deficiency of complete, accurate, and timely data available for decision making. Also, if data is inaccessible, then it may decrease the effectiveness of data users who rely on it for performing tasks.

However, some researchers have found contrasting evidence related to the impact of data quality on performance. For instance, Warth et al. (2011) have found there is no significant effect of data quality on planning performance. Kim et al. (2012) have found the quality of data has no significant effect on innovation. Hazen et al. (2014) recognised the importance of data quality from the analysis of a case study. The authors found that the quality determines the usability of the data and the reliability of insights generated from it. The impact of BDA resource on firms' competitive advantage depends on data and information quality. BDA activity does not create value for the organisation if data and information quality is not attained. Setia et al. (2013) have found out that information/data quality has a significant effect on customer service performance. Information quality is also found to positively influence information satisfaction (Nelson et al. 2005).

Hence, the following hypotheses are proposed:

H9a: The operational performance of an organisation is positively affected by its ability to maintain data and information quality.

H9b: The innovation performance of an organisation is positively affected by its ability to maintain data and information quality.

In addition to the hypothesized direct relationships: 1. between BDA maturity and data and information quality (DIQ) and, 2. between DIQ and operational and innovation performance, it is possible to identify an indirect relationship through DIQ by considering the hierarchy of capabilities perspective and the RBV. The ability to maintain DIQ is an organisational capability, and by examining its role, the relationship between BDA maturity and operational and innovation performance can be defined. Several studies (Haryadi et al. 2016; Pipino et al. 2002; Wang and Strong 1996; Shen et al. 2015; Kwon et al. 2014; Song et al. 2017; Abdullah et al. 2015; Hazen et al. 2014) have recognised that operational and innovation performance is influenced by data quality. Wiengarten et al. (2010) found out that joint decision-making significantly influences operational performance only in the presence of high-quality information. Moreover, based on a simulation experiment, it is found that higher levels of information quality significantly increases the decision quality (Raghunathan 1999; Janssen et al. 2017). Citroen (2011) interviewed chief executives to understand the decision-making process and documented the emphasis put on information quality for strategic-decision making. When asked about the requirement of information quality, a Chief Executive Officer (CEO) has stated that "*Quality of information means integrity, robustness, able to stand up for scrutiny, but very important is also a guarantee of completeness, wholeness*"(Citroen 2011, p.497). Gustavsson and Wänström (2009) explained the importance of information quality via case studies on manufacturing companies. The authors found out that deficiencies in reliability and timeliness of information increase the production lead time and hindered production flexibility; deficiencies of completeness of information resulted in decreased responsiveness to deliver products on time; and deficiencies in reliability and accessibility of information resulted in high inventory cost as the credibility of forecasted information will be low. Moreover, Redman (1998) argued that, at the operational level, increased cost, lack of job satisfaction and customer satisfaction are the consequences of poor data quality. Lee et al. (2006) performed a cost-benefit analysis of data quality and argued that costs of data quality are easily quantifiable but benefits of good quality data are intangible. The problems due to poor data quality cost organisations an average of 8-12 per cent of revenue (Redman 1996). Similarly, Selvage et al. (2017) have found that poor data quality results in the loss of $15 million/year on average. Therefore, it can be noted from the literature that improved data and information quality help to sustain competitive advantage and improve performance through decreasing manufacturing costs, improving customer satisfaction, and eliminating non-value-added activities. Moreover, monitoring supplier quality requires maintenance of a supplier performance database that can provide an accurate track of supplier quality performance data. The availability of quality data about supplier performance can support employees to solve problems such as poor product quality and issues with a delivery that may stem from the supplier side (Krause et al. 1998), and can also enhance innovation (Kim et al. 2012). Data accuracy has also been found to increase planning quality in the manufacturing sector (Chae et al. 2014).

Though several studies have highlighted that operational / innovation performance is dependent on data quality, authors such as Warth et al. (2011) have contended that further investigation is required to estimate the role of data quality in improving performance. This, along with the hypothesized link of BDA maturity and operational and innovation performance, and within the hierarchy of capabilities perspective, suggests an indirect relationship between BDA maturity and operational and innovation performance, mediated by data and information quality. Such an inference is consistent with prior research demonstrating the mediating role of data quality (Sambamurthy et al. 2003; Vickery et al. 2010). In agreement with these theoretical and empirical arguments, the following hypotheses are formulated:

H10: DIQ mediates the relationship between BDA maturity and operational performance.

H11: DIQ mediates the relationship between BDA maturity and innovation performance.

3.3.5 The mediating role of SCA capabilities on the relationship between BDA maturity and operational and innovation performance.

Supply Chain Analytics (SCA) capabilities, also referred to as data-enabled / data-driven supply chain capabilities, is an operational capability that enhances a firm’s ability to coordinate and perform various activities such as sourcing, planning, etc. Unlike dynamic capabilities, operational capabilities provide the drive for a firm “to make a living in the present” (Brusset 2016, p.47). Supply chain managers rely on data-driven insights for various purposes such as to gain visibility, collaboration, process control, monitoring, optimisation, etc. and, ultimately aim to gain competitive advantage (Hazen et al. 2014; Davenport 2006). For example, "Cisco systems has created a digital platform for the near real-time transformation of information between customers, contract manufacturers, and logistics providers. This enables responsive collaborative planning and efficient coordination of resources across its global supply chain" (Rai et al. 2006).

In a supply chain environment, data is stored in a large distributed database or sporadically connected database systems such as mobile or cloud computing systems (Pitoura and Bhargava 1999). Similarly, Rai et al. (2006) have argued that in a supply chain environment, data consistency is a greater problem as the data generated systems are fragmented and spread widely across supply chain boundaries. The usage of BDA to manage supply chain activities will enable firms to automate the system and accurately capture supply chain data. Big data technology-enabled supply chains can enable process integration and improve operational performance (Kache 2015; Chircu et al. 2014). Chae, Yang and Olson (2014) have found that data-enabled supply chain planning has a significant positive effect on planning satisfaction and operational performance (order fulfilment, delivery as promised, delivery flexibility, flexibility to change output volume, flexibility to change product mix). Jamehshooran et al. (2015) assessed the business value of utilising analytics to perform supply chain activities such as planning, sourcing, making, delivering and returning. The authors found that the effect of analytics in supply chain planning had a greater effect on operational performance. An analytical experiment conducted by Zou et al. (2016) revealed that timely/real-time processing of supply chain network data can reduce supply chain risk and improve operational performance. Apart from these, big data can be used in various supply chain functions, such as sourcing, distribution, and networking (Lavalle et al. 2011; Sanders 2016). Sanders (2016) argued that analytics can be used in supplier risk assessment, facility location and layout, scheduling, etc. Sherer (2005) argued that analytics techniques can be used for supply chain mapping and supply chain visualisation. Using real-time supply chain data and in-memory analytics, firms can optimise supply chain processes and manage demand planning to reduce cost, and increase product quality by reducing defects (McAfee and Brynjolfsson 2012; Fawcett and Waller 2014). Data-enabled supply chains enhance collaboration and improve understanding of market demands enabling firms in the supply chain to quickly respond to changing needs (Sanders 2014).

Further, BDA can help optimise supply chain activities by obtaining internal and external data from customers, suppliers, and competitors. SCA capabilities might facilitate firms to process the supply chain data and would offer several of the aforementioned benefits. The external information or knowledge provided by the use of analytics to manage supply chain activities would enhance efficiency and effectiveness of firms and could reduce cost, improve product quality, delivery performance and innovate new products and services. So, this research suggests that SCA as an operational capability is critical to improving operational and innovation performance. Therefore, the following hypotheses have been proposed in this research, which suggests that supply chain analytics capabilities can achieve better operational and innovation performance:

H12a: Supply chain analytics capabilities is positively related to operational performance.

H12b: Supply chain analytics capabilities is positively related to innovation performance.

Moreover, in the literature, there are several studies that support the direct effect of SCA/data-enabled/data-driven supply chain on firm performance (Yu et al. 2017; Gunasekaran et al. 2017; Sangari and Razmi 2015). However, antecedents and consequents of supply chain analytics and its mediating role on the relationship between intra-organisational BDA maturity and firm performance have received limited research attention. Chae, Yang, Olson, et al. (2014) investigated the importance of supply chain capabilities on the relationship between advanced analytics and operational performance. The findings of their study suggest that the direct effect of advanced analytics on operational performance is low, but there is a significant positive effect when it is mediated by supply chain capabilities such as Just-in-time or, statistical process control. These supply chain capabilities are quantitative data-driven approaches which are significant to enhance processes. Similarly, the authors found out that advanced analytics have a positive effect on manufacturing planning quality only when mediated through these data-enabled supply chain capabilities. Based on the above arguments and drawing upon the theory of RBV and hierarchy of capabilities view, it can be suggested that intra-organisational BDA maturity enhances supply chain analytics capabilities, this, in turn, influences operational and innovation performance. Hence, the following hypotheses are proposed depicting the mediating role of SCA capabilities:

H13: Supply chain analytics capabilities mediates the relationship between BDA maturity and operational performance.

H14: Supply chain analytics capabilities mediates the relationship between BDA maturity and innovation performance.

3.4. Summary of theory and hypothesis development

In summary, this chapter has provided a theoretical foundation for developing the conceptual model in support of empirically investigating the phenomenon of BDA practice and value creation and to address the research questions. This chapter combined the views of resource-based theory, dynamic capabilities and the hierarchy of capabilities to explain the potential of BDA in creating value for organisations adopting it. Further, this chapter proposed a number of hypotheses to explain the role of absorptive capacity, data and information quality, and supply chain analytics capability on the relationship between intra-organisational BDA maturity and operational and innovation performance (Table 3.2). In addition, several pieces of evidence from prior studies are considered to reinforce the proposed hypotheses. Using the conceptual framework, the current study explores the relationships with a drive towards theory building.

Table 3.2 List of proposed hypotheses in this researchNo	**Hypothesis**
RQ1a.	**What is the relationship between Big Data Analytics capability maturity and firm performance dimensions?**
H1	Big data analytics maturity has a significant positive effect on operational performance dimensions
	H1a: Big data analytics maturity has a significant positive effect on product quality performance
	H1b: Big data analytics maturity has a significant positive effect on cost performance
	H1c: Big data analytics maturity has a significant positive effect on flexibility performance

	H1d: Big data analytics maturity has a significant positive effect on time-based performance
	H1e: Data Generation capability has a significant positive effect on overall operational performance
	H1f: Data Integration capability has a significant positive effect on overall operational performance
	H1g: Advanced Analytics capability has a significant positive effect on overall operational performance
	H1h: Digital Analytics capability has a significant positive effect on overall operational performance
	H1i: Data Visualisation capability has a significant positive effect on overall operational performance
	H1j: Big Data Skills has a significant positive effect on overall operational performance
	H1k: Data-Driven Culture has a significant positive effect on overall operational performance
H2	Big data analytics maturity has a significant positive effect on innovation performance.
	H2a: Data Generation capability has a significant positive effect on Innovation performance.
	H2b: Data Integration capability has a significant positive effect on Innovation performance.
	H2c: Advanced Analytics capability has a significant positive effect on Innovation performance.
	H2d: Digital Analytics capability has a significant positive effect on Innovation performance.
	H2e: Data Visualisation capability has a significant positive effect on Innovation performance.
	H2f: Big Data Skills has a significant positive effect on Innovation performance.
	H2g: Data-Driven Culture has a significant positive effect on Innovation performance.
H3	Big data analytics maturity has a significant positive effect on Absorptive capacity
H4	Big data analytics maturity has a significant positive effect on data and information quality
H5	Big data analytics maturity has a significant positive effect on supply chain analytics capability
RQ1b.	**What is the role of Absorptive Capacity on the relationship between BDA capability maturity and firm performance?**
H6a	Absorptive capacity is positively related to operational performance.
H6b	Absorptive capacity is positively related to innovation performance.
H7	Absorptive capacity mediates the relationship between BDA maturity and operational performance.
	H7a: Absorptive capacity mediates the positive effect of Big Data Analytics maturity on product quality performance
	H7b: Absorptive capacity mediates the positive effect of Big Data Analytics maturity on cost performance

	H7c: Absorptive capacity mediates the positive effect of Big Data Analytics maturity on flexibility performance
	H7d: Absorptive capacity mediates the positive effect of Big Data Analytics maturity on time-based performance
H8	Absorptive capacity will mediate the relationship between BDA maturity and innovation performance.
RQ1c.	**What is the role of Data and Information Quality (DIQ) on the relationship between BDA capability maturity and firm performance?**
H9a	The operational performance of an organisation is positively affected by its ability to maintain data and information quality.
H9b	The innovation performance of an organisation is positively affected by its ability to maintain data and information quality.
H10	DIQ mediates the relationship between BDA maturity and operational performance.
	H10a: DIQ mediates the positive effect of Big Data Analytics maturity on product quality performance.
	H10b: DIQ mediates the positive effect of Big Data Analytics maturity on cost performance
	H10c: DIQ mediates the positive effect of Big Data Analytics maturity on flexibility performance
	H10d: DIQ mediates the positive effect of Big Data Analytics maturity on time-based performance
H11	DIQ will mediate the relationship between BDA maturity and innovation performance.
RQ1d.	**What is the role of Supply Chain Analytics capability (SCA) on the relationship between BDA capability maturity and firm performance**
H12a	Supply chain analytics capabilities is positively related to operational performance.
H12b	Supply chain analytics capabilities is positively related to innovation performance.
H13	Supply chain analytics capabilities mediates the relationship between BDA maturity and operational performance.
	H13a: SCA mediates the positive effect of Big Data Analytics maturity on product quality performance.
	H13b: SCA mediates the positive effect of Big Data Analytics maturity on cost-based operational performance.
	H13c: SCA mediates the positive effect of Big Data Analytics maturity on flexibility performance.
	H13d: SCA mediates the positive effect of Big Data Analytics maturity on time-based performance.
H14	Supply chain analytics capabilities will mediate the relationship between BDA maturity and innovation performance.

Chapter 4 Research Methodology

4.1 Chapter introduction

This chapter provides information on the research methodology adopted to test the hypotheses related to BDA capabilities on firm performance and the method adopted to explore the phenomenon of the digital divide. First, this section discusses and justifies the philosophical standpoint, research strategy and the choice of data collection method adopted in this study. Then, it provides a justification for utilising factor analysis and Structural Equation Modelling (SEM) techniques to test the hypotheses. Further, the rationale behind the use of cluster analysis to investigate the digital divide is discussed.

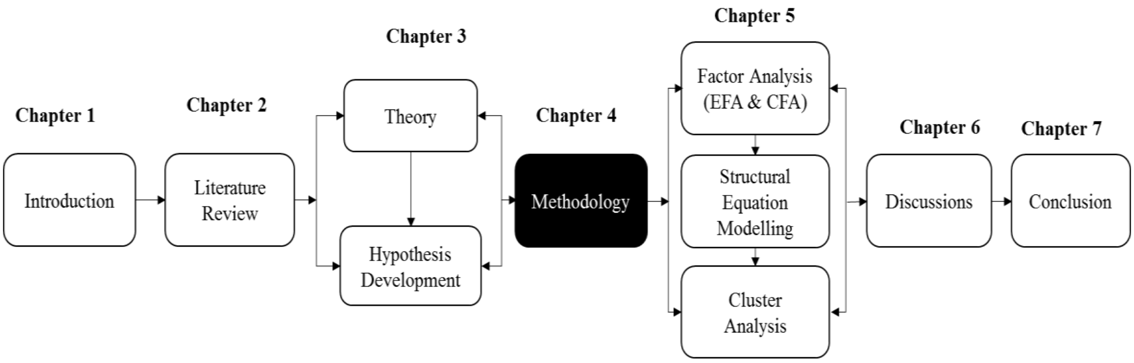


Figure 4.1 Position of the chapter in this thesis

4.2 Methodology

The main objective of identifying an appropriate research methodology is to ensure the internal and external validity of the study. Various aspects of the research such as the quality of the research design, the ability of the chosen approach to validate the research model, accuracy and the relevance of the data, replicability and generalisability are some of the key criteria considered to ensure the internal and external validity (Straub 1989). The selection of a suitable approach occurs after the identification of research questions, the variables of concern, and the specification of the model to be assessed empirically. Scholarly works on research methodology illustrate that there are various types of research methodology commonly used in social science. The current research is interdisciplinary in nature combining social science and information systems. Ever since its inception, the Information Systems (IS) field has employed a variety of methods to address specific issues related to the discipline (Robey 1996). Together with emphasising the necessity of methodological pluralism in the IS discipline, Robey (1996) argued that a suitable theoretical foundation and research methods can be chosen as long as it is justified by the research aim on realistic grounds (Figure 4.2). Consequently, as shown in Figure 4.3, the design of this research is based on both philosophical and practical considerations. A detailed discussion of these aspects of the research design is given in the following sections.

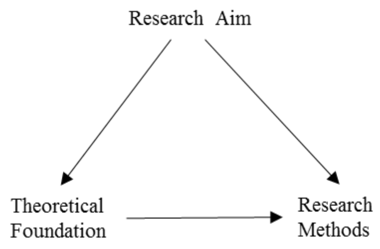


Figure 4.2 Triad for the justification of research

Source: (Robey 1996)

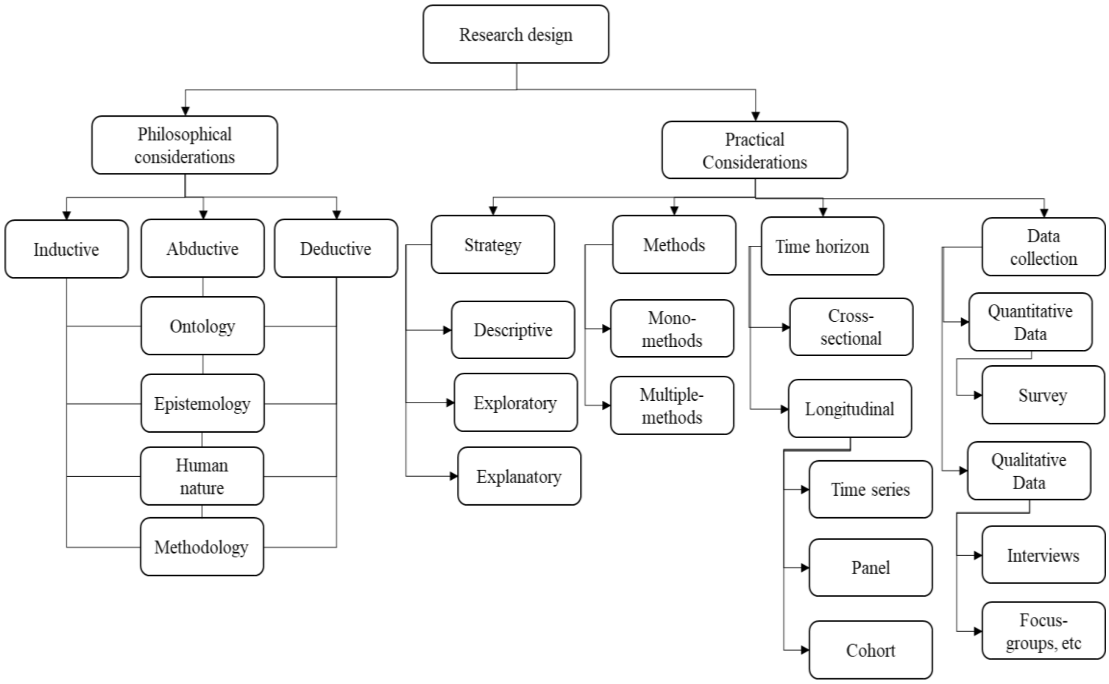


Figure 4.3 Research design considerations

Source: adapted from (Saunders et al. 2016) (Bryman and Bell 2011) (Nassar 2011) (Neuman 2014)

4.3 Philosophical considerations

Social science research, in general, depends on the philosophical inclination of the researcher, which in turn underpins research strategy and research methods. The research paradigm denotes the world-view and beliefs about knowledge and includes elements such as ontology, epistemology, human nature (Burrell and Morgan 1980), methodology (Denzin and Lincoln, 1994, as cited in Frankel et al. 2005) and axiology (Saunders et al. 2016). Considering the interactive nature of the phenomenon and the inquirer, it is argued that in an ideal situation the inquirer should not influence the phenomenon, and vice-versa (Guba and Lincoln 1994). So, it is necessary to decide on an appropriate research paradigm before going further to select methodological approaches (Goles and Hirschheim 2000; Guba and Lincoln 1994). Moreover, Frankel et al. (2005) argued that research paradigm and research methods are two different aspects and it is the choice of a research paradigm that determines the research method suitable for the researcher. Burrell and Morgan (1980) have claimed that social scientists view their subjects of interest through explicit or implicit assumptions about the nature of the social world and how it can be investigated. Accordingly, the four sets of assumptions: ontological nature, epistemological nature, human nature (which provides implications for the last one), and methodological assumptions are thoughtfully considered in this research, as they lay the foundation for the research methodology.

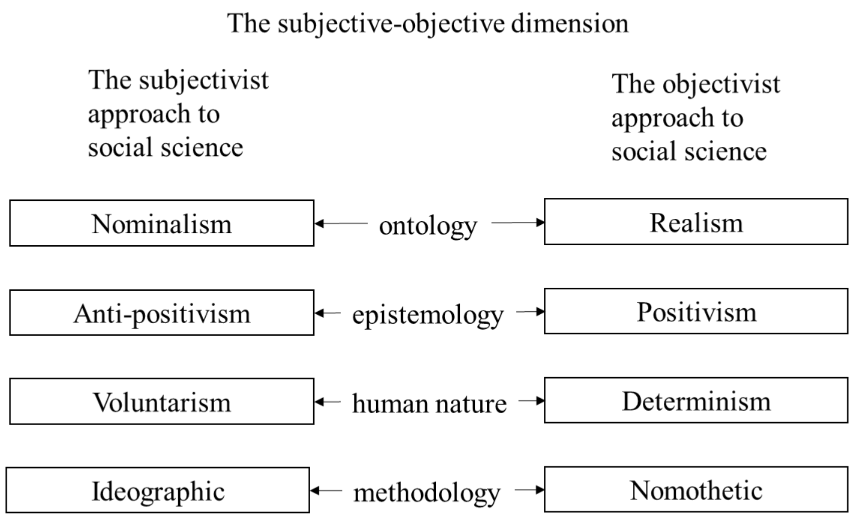


Figure 4.4 A scheme for studying philosophical assumptions about the nature of social science

Source:(Burrell and Morgan 1980)

4.3.1 Ontological nature

As stated in Burrell and Morgan (1980, p. 2), researchers are faced with some basic questions on ontology; “whether the 'reality' to be investigated is external to the individual - imposing itself on individual consciousness from without - or the product of individual consciousness; whether 'reality' is of an 'objective' nature or the product of individual cognition; whether 'reality' is a given 'out there' in the world, or the product of one's mind”. According to Saunders et al. (2016), ontology is concerned about the nature of reality, and Bryman and Bell (2011, p.20) argued that ontology can be referred to as "the nature of social entities", and social entities can either be considered objectively as "reality external to social actors", or subjectively constructed inherent to the "perceptions and actions of social actors". Moreover, Frankel et al. (2005, p.186), stated that ontology "includes claims about what exists, what it looks like, what units make it up and how these units interact with each other". Ontological assumptions determine how the researcher perceives the world, and the researchers' ontological position can be identified within the continuum of objectivism to subjectivism or constructionism. Objectivism embraces realism and perceives social actors and social entities as being independent, and the objects can be measured in a similar way to the approaches of natural science research (Saunders et al. 2016). Whereas, subjectivists embrace nominalism and believe that the existence of the social phenomenon is dependent on social actors. Nominalists do not acknowledge the existence of "'real' structure to the world", but realists believe in "tangible and immutable structures" (Burrell and Morgan 1980, p.4).

4.3.2 Epistemological nature

In contrast to ontology, epistemology is concerned with "assumptions about knowledge, what constitutes acceptable, valid and legitimate knowledge, and how we can communicate knowledge to others" (Saunders et al. 2016, p.127). As shown in Figure 4.4, the assumptions of epistemology are generally identified as positivism and anti-positivism (Burrell and Morgan 1980). According to Saunders et al. (2016), the epistemological stand of positivism and anti-positivism are varied in several ways. While the former assumption supports the usage of natural science methods to social science, the latter assumption supports the usage of methods originated from arts and humanities. Positivism is also referred to as objectivism, which embraces 'realism' ontology. Whereas, anti-positivism is also referred to as subjectivism/interpretivism - a binary opposite of positivism, which embraces 'nominalism' ontology. Objectivists consider facts and numbers as acceptable knowledge with a notion that the phenomenon is readily observable. Whereas, subjectivists view narratives and opinions generated through social interactions as an acceptable knowledge with an assumption that knowledge is socially constructed. The way anti-positivism sees the social world is 'relativistic' and can be understood only from the perspective of individuals involved in the activities under investigation. Anti-positivists reject the idea of ‘observer’ and the social world must be understood from the inside by being part of it rather than from outside as an observer.

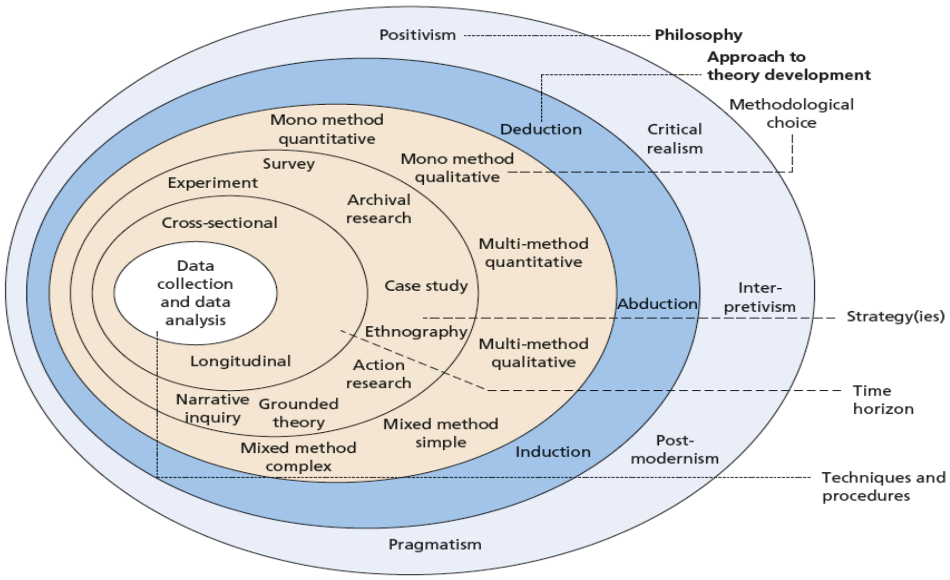
4.3.3 Human nature

The debate on the ‘human nature’ paradigm is concerned with the relationship between humans and the environment. It is related to the way humans respond to external forces. Two assumptions under this category are *deterministic* and *voluntarism*. While the former perceives that humans are conditioned and controlled by their external circumstance, the latter sees that humans are the creator of the environment. According to Neuman (2014), the model of human nature emphasised in positivistic social science research is deterministic, i.e. people respond like ‘robots’. The model of human nature emphasised in interpretivist/subjectivist social research is voluntarism, i.e. humans are autonomous and free-willed. However, it is argued that social scientists tend to choose a middle ground when ‘human nature’ is concerned (Burrell and Morgan 1980). So, in social science research, researchers should take an intermediate perspective accounting for the influence of situational and voluntary aspects on human beings.

4.3.4 Methodological nature

The debate on methodological assumptions “provides a rationale for the ways in which researchers conduct research activities” (Briggs et al. 2012). Burrell and Morgan (1980) have differentiated the methodological assumptions into ‘nomothetic’ and ‘ideographic’. According to the ideographic approach, to understand the social world the researcher should concentrate their efforts on getting close to the subject of interest and explore its detailed background. During the investigation process, the ideographic method allows subjects to unfold its characteristics. Whereas, the nomothetic approach to social science research places emphasis on conducting research using systematic techniques and protocols, and relies greatly on causal laws and law-like statements (Neuman 2014; Gill and Johnson 2002). Testing hypothesis with scientific rigour using quantitative techniques such as surveys and scientific instruments is prominent tools used in nomothetic approach. Eventually, the nomothetic approach is epitomised in positivism and ideographic is associated with ‘anti-positivism’, as illustrated in Figure 4.4.

Apart from the various paradigms discussed earlier, Saunders et al. (2016) also mention axiology. Unlike ontology and epistemology, axiology deals with the social aspects of the research process. The values and ethics associated with conducting a research are the prime concerns of axiology. These assumptions provide guidance to increase the consciousness of the subject of the investigation. It allows the researcher to be aware of the different forms of knowledge that can be obtained and how that knowledge can be categorised into a true knowledge or not. Moreover, the nature of knowledge can either be hard and tangible or soft, subjective, and of the transcendental kind.



Source: (Saunders et al. 2016)

Figure 4.5 The research onion

As shown in Figure 4.5, there are various philosophical assumptions established such as positivism, critical realism, interpretivism, and pragmatism drawing explicitly or implicitly upon a set of ontological beliefs and epistemological assumptions (Briggs et al. 2012). So, depending on the researcher's understanding of the phenomenon under investigation and orientation on ontological and epistemological assumptions, a philosophical position suitable for this research is chosen. A detailed discussion is given in the following section, which help to decide the methodological approach adopted in this study.

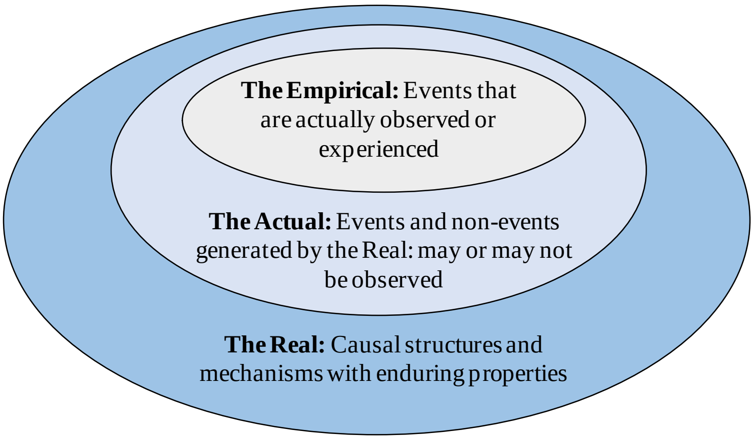
4.4 Philosophical assumptions–critical realism

In this section, the logic behind some of the major philosophical stands such as positivism, interpretivism, critical realism, and pragmatism are considered. The rationale for choosing critical realism as a philosophical choice of this research is also discussed. This choice is based on the ontological and epistemological assumptions debated earlier.

Ontologically, positivists see the social world as a concrete structure, but interpretivists perceive the social world as a concrete process (Evely et al. 2008). Positivists assume that the social world exists similar to a laboratory setting, within a controlled environment, facilitating data collection (Edwards et al. 2014). Positivists are interested mainly in examining observable objects and are closely associated with quantitative research design. However, Laing (1967, p. 53) as cited in (Gill and Johnson 2002), has discussed some of the major criticisms of positivism: (1) compared to the subject-matter of natural sciences (physical objects), the subject-matter of the social sciences are humans, and the causal model approach is criticised as human actions contain internal logics which are the main interests of social science, (2) the natural science does not imply the subjective element of human beings, so its methodology is inappropriate for social science, (3) the social world cannot be comprehended just by investigating a causal relationship, as a lot depends on the context, and sound understanding of the social phenomenon is possible only if the subjective dimensions are considered.

Similarly, anti-positivism/interpretivism has also been criticised for several reasons (Briggs et al. 2012). The first criticism is that, as the reality is multi-perspectival, the way humans create meaning of reality is affected by contextual biases. Second, since human behaviour is a routine process, it is relatively unusual for humans to reflect on their behaviour in a structured manner. Third, since participants of the research are unaware of the broader structures that govern their interpretation, their accounts of themselves, of others and of events can be incomplete. So, to capitalise on the strengths of both positivism and interpretivism and overcome weaknesses of it, some scholars advocate the use of pluralistic paradigms (Bryman and Bell 2011; Creswell 2014; Neuman 2014). On a different spectrum, a pragmatism world view does not commit to any assumptions about reality, but rather focuses on the research problem and encourages the use of all available approaches (multiple methods) to understand the problem. Pragmatists have the freedom of choice in terms of "methods, techniques, and procedures of research that best meet their needs and purposes" (Creswell 2014).

Extreme positivism and extreme subjectivism are two contrasting philosophies widely applied in the social science research, but other philosophical approaches exists between this continuum, and critical realism is one among them (Evely et al. 2008). In contrary to extreme positivism and interpretivism, the critical realism (CR) philosophy provides an alternate perspective to conduct social science research. As stated in Evely et al. (2008), critical realism denotes that "there are objectively knowable, mind-independent realities, but the influence of human perception and cognition in shaping that reality is acknowledged". Similarly, Easton (2010) stated that "Critical realism assumes a transcendental realist ontology, an eclectic realist/interpretivist epistemology, and a generally emancipatory axiology". The proponents of critical realism claim that the reality should be critically examined in order to apprehend it, but also acknowledge that apprehending reality are always imperfect (Guba and Lincoln, 1994). Moreover, CR considers that ontology is stratified into real, empirical and actual, and emphasis on the underlying causes of the phenomenon under investigation (see Figure 4.6) (Edwards et al. 2014).



(Saunders et al. 2016)

Figure 4.6 Stratified ontology of critical realism

In general, CR accepts the existence of objective reality and believes in causal language, but equal importance is given to the influence of cognitive reasoning in shaping the truth, i.e. causal language with thinking (Easton, 2010). Edwards et al. (2014) argued that in contrast to CR philosophy, positivists fail to identify the mechanism through which the two phenomena are related. The comparison between major philosophical approaches discussed here is given in Table 4.1.

Table 4.1 Comparison between major research philosophies

	Positivism	Critical Realism	Interpretivism	Pragmatism
Ontology (nature of reality or being	• Real, independent and external to social entities. • Single reality	• Real, but imperfectly. • External, Independent and intransient. • Stratified ontology (empirical, actual, and real). • Causal mechanism.	• Nominal, constructionist ontology. • Multiple Realities.	Reality is complex and external and exists because of ideas.
Epistemology (what constitutes acceptable knowledge) Axiology	• Facts are observable and measurable. • Causal explanation and prediction as contribution	• Epistemological relativist. • Historical causal explanation as contribution. • Knowledge is obtained by uncovering causal mechanism.	• Focus on narratives, stories, perceptions and interpretation. • Avoids quantitative analysis.	• The practical meaning of knowledge in specific contexts • 'True' theories and knowledge are those that enable successful action
Axiology (role of values) Typical	The researcher is detached and truly maintains objective stands to conduct value-free research	Research is value-laden, and researcher tries to be objective and acknowledges bias and errors in order to minimize it.	Research is value-bound, and the researcher is truly subjective, and interpretations are key to knowledge creation.	The research itself is initiated towards answering researchers' doubts and queries about a phenomenon and conducted in a value-driven approach.

Source: adapted (Saunders et al. 2016)

This study aims to observe the impact of BDA capabilities on firm performance, through uncovering causal mechanisms. The conceptual model developed to explain the relationship contains constructs such as BDA capabilities and

ACAP. In Saunders et al. (2016), it is argued that positivist approach is inadequate and hinders the ability of the researcher to thoroughly examine the relationships between unobservable objects and to give importance to the social interactions. In this research, the phenomenon under investigation is viewed from the socio-technical perspective to understand the reality. In case of this research, people are the decision makers, who interact with 'technology' which is BDA in order to improve 'processes'. Furthermore, Fleetwood (2004) argued that CR differentiates real entities into materially real, ideally real, artefactually real and socially real. There is no clear distinction, but they tend to overlap: 1) *artefactually real* refers to entities such as technologies, computers, hole in the ozone layer, etc., 2) *materially real* refers to natural phenomenon such as mountains, ozone layer, etc., 3) *ideally real* refers to discourse and discursive entities, and 4) *socially real* refers to practices, social structures, etc. which cannot be recognised with touch or smell and its existence is dependent on human activity.

Drawing from these thoughts, BDA technology can be considered as an artefactually real entity and BDA practice by organisations can be considered as socially real entities, while both constantly interact with each other within a social system. Therefore, the researcher views reality from a critical realist perspective, and inclined towards assumptions about knowledge as an external entity, independent of the researcher, and believe in the existence of measurable truth within social entities. However, the researcher also recognises the importance of context and the influence of cognitive reasoning in shaping the reality and extracting knowledge from it.

4.5 Research approach

Similar to the philosophical stands, theoretical stands of the research are also significant to make an informed decision about the research design (Saunders et al. 2016). Different approaches such as deduction, induction, and abduction are used to deal with issues concerning theory development. Deductive reasoning approach starts with theory and data is collected to test the hypothesis developed from the theory. Whereas inductive reasoning starts with data collection to explore a phenomenon and conceptual framework is developed based on the themes and patterns identified from the data. Deductive approach concludes with verification of existing theory, but a new theory is generated as an outcome of inductive reasoning. In contrast, abduction combines both inductive and deductive approaches and moves back and forth from incomplete observation to gain the best possible explanation (Saunders et al. 2016). This research takes an approach of deductive reasoning aiming towards theory building. Accordingly, in this research, a suitable methodology is adapted to test the hypotheses.

Table 4.2 Research approach and its meaning

Approaches	Meaning
Induction	leading into
Deduction	leading to separation, removal, or negation
Abduction	leading away from
Retroduction	deliberately leading backward

Source: (Chiasson 2005)

4.6 Practical considerations

4.6.1 Research strategy

Research can be classified into various types based on the research purpose, outcome and data. For instance, based on the research purpose, Yin (2008) has identified three major types of research; descriptive, explanatory and exploratory. Moreover, grounded on the anticipated research outcome, it is classified into basic or pure research and applied research (Hedrick et al. 1993; Saunders et al. 2016; Neuman 2014). On the other hand, Creswell (2014) classified the research into qualitative, quantitative and mixed methods based on the type of data used in the research. The strategic position of this research is depicted in Figure 4.7.

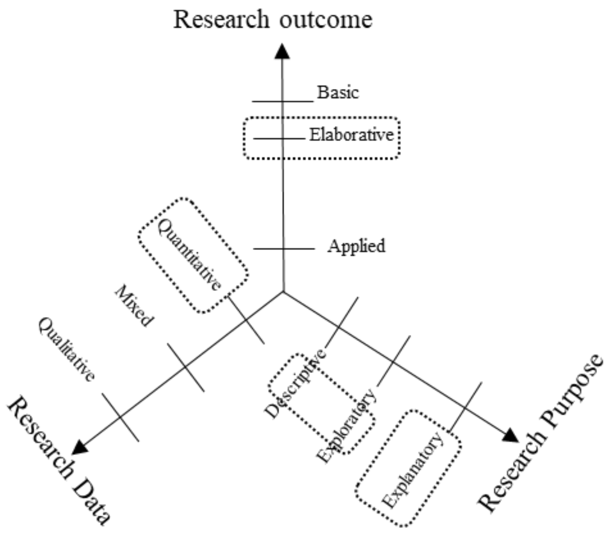


Figure 4.7 Research positioning

In terms of the research purpose, this research is positioned predominantly as explanatory and descriptive in nature, though it has a slight inclination towards exploratory. Exploratory research is conducted only when a very little is known about the phenomenon and, meant to offer scope for further examination of the phenomenon (Yin 2008). Explanatory, on the other hand, aims to establish a causal relationship between variables. In explanatory research, quantitative data can be collected to perform statistical tests such as regression to explain ‘How’ a relationship occurs between the variables of interest about a phenomenon, or use qualitative data to explain ‘why’ a cause and effect relationship occurs (Saunders et al. 2016). Since the aim of this research is to explain the causal relationship between variables illustrated in the research model, explanatory research is adopted in this study.

However, in the case of the second aim of this research (to understand the phenomenon of the digital divide), very little is known about it. The most suitable strategy will be a comprehensive exploratory research conducted by ways such as a literature search, in-depth interviews and focus groups. Adopting an exploratory research strategy to address this research aim would provide greater insights into the phenomenon. Yet in this research, a descriptive-exploratory strategy is adopted since, in practice, descriptive and exploratory research often blur together (Neuman 2014). Moreover, descriptive research allows a researcher to “gain an accurate profile of events, persons or situations” (Saunders et al. 2016, p.175), or “create a set of categories or classify types” (Neuman 2014, p.38). In this study, the aim is to categorise small, medium and large organisations based on the BDA capabilities maturity level and profile their current situation/state, thereby exploring the potential divide between them.

Moreover, based on the understanding of the phenomenon of BDA practice on value creation and the phenomenon of the digital divide, it is argued that the measurement of this phenomenon is achievable via quantitative data collection. Hence, in terms of research data, this study is positioned as simply quantitative.

Finally, when it comes to the outcome of the research, ‘basic research’ is intended to expand knowledge (i.e. knowledge is the key motivator, although the outcome of basic research will eventually help address specific problems it’s not the driving goal) by discovering statistically significant relationships (Hedrick et al. 1993). Whereas, applied research is intended to improve the understanding of a specific problem, which is the motivating factor. Although basis and applied research differ in several aspects, Hedrick et al. (1993) argued that they fall on a continuum. This implies that based on research outcome, it can be classified into the range between basic to applied. Based on these arguments, it can be argued that this research is in between basic and applied, as it expands the knowledge of BDA practice and as well as helps in understanding the specific problem of digital divide.

4.6.2 Research methods

Research methods are, in general, the data collection or fact-finding procedures which can produce insights about the phenomenon under research. The method of data collection and the measures used to capture it are also dependent on the theoretical framework, research questions, and objective (Frankel et al. 2005). Some researchers classify the research methods based on the nature of data into qualitative and quantitative research methods. However, there are two broad categories of research methods described in (Saunders et al. 2016); Mono and multiple methods. Quantitative or qualitative studies are the two options available under the mono-method approach. However, there are several categories of multiple methods such as multi-method quantitative, multi-method qualitative and mixed methods.

Based on the work of Easterby-Smith et al. (2002), Frankel et al. (2005) have summarised the dominant research methods used in the SCM and logistics studies. Frankel et al. (2005) found that most researchers in the SCM discipline have adopted mono-methods research. Even in BDA domain, there are existing academic literature (Cao et al. 2015; Wang et al. 2016a) that have adopted purely quantitative research design. Similarly, this research uses the Mono-method (quantitative) approach. Moreover, the rationale for choosing mono-method quantitative research design is supported by the fact that for both testing of causal relationships between variables in the research model and to understand the current state of BDA maturity of SMEs and large organisations, an approach that is unbiased and as objective as possible is required. It is not practically convenient to attain these aims using a mono-method qualitative research design, as it requires several stages of data collection and associated with potential biases.

In terms of the time dimension, research can be performed either cross-sectional or longitudinal (Neuman 2014). Longitudinal research is used to capture social processes or changes that require data gathering at different time intervals. This study is interested only in capturing a 'snapshot' of current BDA practices and testing its effect on various organisational and performance factors. Moreover, the researcher does not intend to test the influence of BDA practices across different time periods. Hence, the cross-sectional approach is considered as the most suitable option. Consequently, this study has adopted a mono-method approach for quantitative data collection using a cross-sectional survey. A detailed discussion on the rationale behind the application of survey technique and the strategies adopted in this research to design, develop and distribute questionnaire survey is provided in the section 4.10.

4.7 Survey research

Based on both philosophical and practical considerations, this research has adopted a quantitative approach utilising a survey methodology. "A survey design provides a quantitative or numeric description of trends, attitudes, or opinions of a population by studying a sample of that population" (Creswell 2014, p.145). Surveys can be beneficial to examine attitudes and beliefs, explore relationships and gather sensitive information (Colton and Covert 2007; Aldridge and Levine 2001). It is argued that the fundamental principle of the survey process is 'by describing the characteristics of people responded to a survey, it is easy to describe the target population' (Fowler Jr 2014). Similarly, Bryman and Bell (2011) have mentioned the two characteristics of surveys: 1. Surveys use samples to allow for generalisation of a specific concept/relationship within a population, 2. Surveys use a systematic approach for gathering information using instruments such as structured questionnaires or interviews. Survey strategy is often associated with the deductive approach (Saunders et al. 2016), which is in concordance with this research as well.

Gill and Johnson (2002) argued that the survey strategy can be positioned in between experimental and ethnography research; it can take different forms based on the researcher's choice of the investigation. Researchers have recognised the existence of three basic forms of survey research: 1). descriptive, 2). confirmatory (theory testing) and 3). exploratory survey (Malhotra and Grover 1998; Pinsonneault and Kraemer 1993). However, Zikmund et al. (2010) argued that while most surveys are descriptive, they are also often used to explore or explain concepts. Moreover, survey strategy would allow the researcher to generate quantitative data enabling them to perform both descriptive and inferential statistics (Saunders et al. 2016). In general, surveys are regarded as a less costly and more accurate option than other approaches

(Forza, 2002). It can be argued that the most important aspect of the conducting survey research is understanding the target population. The target population of this research is recognised as UK manufacturing companies. The rationale for choosing companies located in the UK and manufacturing as a context is given in section 1.4.

Exploratory elements of the survey: According to Forza (2002), this type of survey research is conducted in the early stages of a study. The main purpose is to gain preliminary insights about a topic which would lead to developing an in-depth survey for future investigation. This approach can help determine the key concepts that need to be measured relevant to the phenomenon of interest. As mentioned previously, some element of exploratory research is involved in this study. Accordingly, in this research, the questionnaire survey is designed to measure the dimensions of BDA capabilities holistically to understand the maturity level of the UK manufacturing organisations. Here, the phenomenon of interest is the relevance of the 'digital divide' to BDA capabilities. The main motive is to gain preliminary insights into the phenomenon by comprehending whether there is a digital divide or not, which will help the researcher to conduct an in-depth investigation of the topic of interest.

Explanatory or confirmatory elements of the survey: This type of survey is conducted to collect data with an aim to test concepts, models and propositions developed using theory about a phenomenon (Forza 2002). Explanatory survey research forms the main basis of this study as one of the aims is to understand the causal relationship between latent variables such a BDA maturity, ACAP, DIQ, SCA capability, and operational and innovation performance.

Descriptive elements of the survey: Surveys can also be descriptive in nature, the response of people to specific questions such as age, gender etc. can be used to describe the characteristics of the sample. The survey instrument used in this research contains various descriptive elements to characterise the respondents.

Although there are several benefits of survey research, this research has acknowledged the limitations as well. Most of the issues related to survey are due to biased samples (Collis and Hussey 2013), lack of goodwill and patience of respondents (Saunders et al. 2016), and sampling errors (Dillman et al. 2014). Moreover, survey research limits the ability to capture the contextual information of the respondents to the fullest, due to exceeding the length of the questionnaire

(Saunders et al. 2016). This study has attempted to overcome these challenges by adopting a systematic approach by following the systematic steps of survey research discussed in Bryman (2012), given in Figure 4.8.

Moreover, in terms of administration of the questionnaire, Bryman (2012) highlighted the two modes of survey administration: 1. structured interview and 2. self-completion questionnaire. The advantages and disadvantages of various types of survey administration are consolidated based on Rea and Parker (2014) and given in

Table 4.3. Based on the merits and demerits of various survey administration techniques, a web-based self-completion questionnaire survey is found to be more suitable and hence adopted in this study. Moreover, several researchers have compared the effectiveness of survey administration techniques such as mail-out and web-based surveys. Cobanoglu et al. (2001) compared the fax survey, mail-out and web-survey using criteria such as response rate, cost and response time. It is found that the web-based and fax survey are more effective in terms of response time, on average 5.97 days and 4 days respectively. But, the response time of the mail-out survey is comparatively high with 16.46 days. In terms of response rate, web-based (44.2%) achieved desirable results compared to mail-out (26.27%) and fax-based survey (17%). Moreover, concerning the quality of web-based survey, it is found to be influenced by gender, age, socio-economic status, and geographic regions (Gosling et al. 2004). Gnambs and Kaspar (2015) have found that compared to the paper-based survey, computerised surveys produce less misreporting of responses. Also, a web-based questionnaire survey is found to produce less unanswered questions and more consistent responses compared to mail-out questionnaires (Rada and Domínguez-Álvarez 2014). Similarly, Ramsey et al. (2016) found that, if clear instructions are provided about the questions and the nature of the study, web-based surveys have the advantage of increasing attention respondents give to the questionnaire. Hence, based on the merits and demerits of various survey administration techniques, a web-based self-completion questionnaire survey is found to be more suitable and hence adopted in this study.

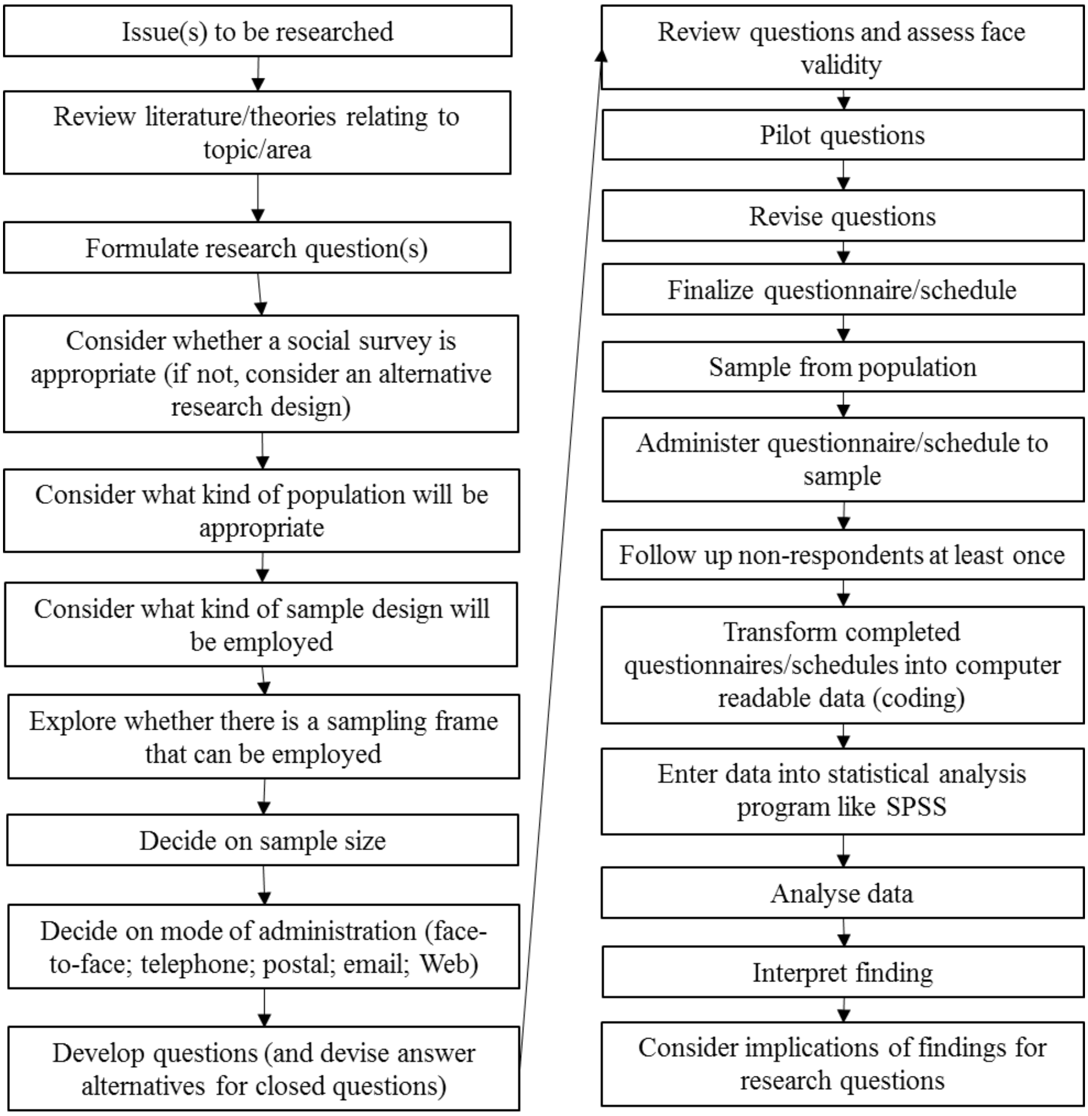


Figure 4.8 Steps involved in survey research

Source: adapted from (Bryman 2012)

Table 4.3 Benefits and drawbacks of different survey administration types

Survey type	Description	Benefits	Drawbacks
Mail-Out Surveys	It involves the distribution of printed questionnaires through the mail using postal services.	1)Less expensive compared to other options such as telephone survey and in-person interviews. 2)It allows respondents to complete it at their convenient time. 3)Respondents have less time constrained and they could consult with others or refer the record to provide more accurate answers.	1)It is time-consuming as it generally takes more time to receive the questionnaire back. 2)Low response rate than telephone surveys. 3)This type of distribution of

		4)It reduces interviewer-induced bias as the structured questionnaire is used. 5)The length and complexity of the questionnaire is not an issue in case of mail-out surveys.	survey is less attractive to respondents poorly educated and with reading deficiencies.
Web-Based Surveys	It is considered as an alternative to the mail-based survey. Respondents are contacted via email and requested to participate in the survey.	1) Respondents can receive the questionnaire and complete it in privacy. 2) It allows rapid and timely collection of information from respondents from the diverse geographical location. 3) It is more cost-effective than a mail-out survey. 4) Similar to the mail-out survey, respondents have less time constrained and they could consult with others or refer the record to provide more accurate answers. 5) Ease of follow-up using email reminders. 6) Security and confidentiality of the information shared by the respondents can be maintained effectively as information is stored on a secure server. 7) It allows the researcher to target specialised population, whose email addresses are known. 8) Web-survey allows the researcher to design questionnaire with 'must answer' option for important questions. This avoids the occurrence of missing data. 9) Like a mail-out survey, it allows the addition of visual aids to enhance respondent understanding.	1) The major drawback is - surveys can only be distributed to the population having email access. 2) It assumes that the respondents have minimum computer literacy. 3) It can increase self-selection bias, and respondents who are not comfortable with web-based systems will opt out themselves. 4) Due to lack of involvement by the interviewer, it is not possible to explain to the respondents if there are any unclear questions.
Web Panels	Panel or web company recruits and incentives participants to participate in various online surveys.	The benefits are the same as a web-based survey, expect that these types of survey are more useful for receiving descriptive feedback on new product and services.	Self-selection bias and limited to people with internet access.

Telephone Survey	Information is gathered from the respondents by trained interviewers using the telephone.	Rapid data collection, less expensive and more anonymity than in-person interviews. Unlike web-based and mail-out surveys, the interviewer can make sure the respondent clearly understands the questions.	Interviewer lack control over the process as respondents may opt out of the interview by simply hanging up the telephone. Less credibility, lack of visual aids and not suitable for complex questions are the other drawbacks of the telephone survey.
In-Person Interviews	Information gathered from the respondents by the interview in-person or face to face	It enables the interviewer to explain unclear questions to the respondents, which will increase the reliability of the information. This technique is more suitable for collecting data from respondents such as criminal offenders on sensitive topics. This technique is highly suitable to reach respondents who do not have internet or telephone accessibility.	The unintentional reaction of the interviewer could affect the future responses by the interviewee and can negatively influence the validity of the questionnaire. Reluctance to co-operate, high stress and less anonymity are some other drawbacks to this method.
Intercept Surveys	This is a variant of the in-person interview survey. Information is gathered from the respondents as they pass by in a public place such as shopping mall and train station.	Cost effective compared to mail-out, telephone, and in-person survey. This is often used as a pre-survey to inform the preparation for the larger survey using telephone, web-based, etc. Similar to the in-person survey, interviewer involvement can make sure respondents understand the questions and also helps to observe the personal characteristics of the respondents.	Lack of anonymity and interviewer bias are some important drawbacks to this technique.

Source: (Rea and Parker, 2014)

4.8 Operationalisation of the constructs

This section will describe how the constructs presented in the conceptual model are defined and measured. The essential element of the survey process is using questions as measures (Fowler Jr 2014). Operationalisation is the process of converting the concepts into questions (Saris and Gallhofer 2014). Accordingly, a researcher should first select a subject and determine the dimensions on which the subject can be evaluated with questions.

From the conceptual model, it is evident that the survey instrument should contain measures related to several constructs (Figure 3.6). In the model, BDA maturity is the independent variable. The three mediating variables are ACAP, DIQ, and SCA capability. There are five endogenous variables, innovation performance and four other latent constructs related to the operational performance dimension (product quality, cost, flexibility and time) which are related to the operational performance dimension. Moreover, the dimension of the BDA Maturity construct is conceptualised as a second-order reflective construct composed of 7 first-order constructs such as data generation capability, data integration and management capability, advanced analytics capability, digital analytics capability, data visualisation capability, data-driven culture and big data skills. Except for BDA Maturity, the measures for ACAP, DIQ, SCA capability and operational and innovation performance are completely derived from the existing literature, mostly from a single study, as the validity and reliability of these constructs have been verified. However, in the case of the dimensions of BDA capabilities constructs, the questions are adapted from multiple sources. The rationale behind the choice of the questions used to measure each latent construct in the model is discussed in the following sections. The structured questionnaire used in this study is developed using the following stages: a) Identification of validated questions from existing literature, b) development of a draft questionnaire, c) initial review of the draft questionnaire by academics, d) pilot test with academic and industry expert, e) refining and finalising the questionnaire.

4.8.1 Operationalising the dimensions of BDA capabilities maturity dimensions.

It is emphasised that variables intended to be measured must be well-defined based on literature review. This strategy is well established and has been applied by researchers in the supply chain discipline (Gligor and Holcomb, 2012). The scales for the BDA Maturity dimensions are proposed in this research based on the suggestions of Churchil (1979), Hinkin et al. (1997), DeVellis (1991) and Mackenzie et al. (2011). Table 4.4 presents an overview of the 7 key capabilities of BDA Maturity and the sources used to develop the measures.

Table 4.4 Scales for BDA capabilities construct

Dimensions	References	No of items
Data generation capability	(Gupta and George 2016) (Hu et al. 2014), (Radcliffe, 2014), (Janssen et al. 2014), (Zhang et al. 2013)	3
Data integration and management capability	(Gupta and George 2016), (Popovič et al. 2012), (Spruit and Sacu, 2015), (Lavalle et al. 2010), (Radcliffe, 2014), (Halper and Krishnan, 2014), (Sulaiman et al. 2015) (Cosic et al. 2015) (Warth et al. 2011)	6
Advanced analytics capability	Multiple sources and developed from the literature review.	5
Digital Analytics	(Cao et al. 2015)	4
Data visualisation capability	(Popovič et al. 2012), (Radcliffe, 2014), (Sulaiman et al. 2015) (Wang et al. 2016b)	4
Data-driven culture	(Gupta and George 2016)	5
Big data Skills	(Gupta and George 2016)	4

4.8.1.1 Items for the data generation capability construct

The data generation capability construct aims to measure the organisation's ability to generate data, and so it determines whether an organisation is data rich or data poor. Data Generation (DG) capability is defined as "the ability of organisations to seek, identify, create, and access data from heterogeneous data sources across organisational boundaries" (Arunachalam et al. 2016). The measures for this construct are developed and adopted from multiple sources such as (Gupta and George 2016), (Hu et al. 2014), (Radcliffe, 2014), (Janssen et al. 2014) and (Zhang et al. 2013). Big data is characterised by its volume, variety, and velocity. The measures used in this research reflect the capability of the organisations in terms of generating a large volume of structured, unstructured and real-time data. Therefore, the respondents are asked to express their level of agreement or disagreement on the following statements regarding DG capability of their organisation: (a) We are able to generate/access/collect a large volume of data from our operations, (b) We are able to generate/access/collect unstructured data (such as textual data, audio and visual data, images, sensor data, RFID data and social media data), (c) We are able to generate/access/collect data in real-time, (d) We are able to generate/access/collect data from a wide range of data sources, and, (e) We have access to very large, unstructured, or real-time data for analysis.

4.8.1.2 Items for the data integration and management capability construct

The data generated by an organisation from various sources need to be integrated to achieve consistency, and this construct aims to measure the ability of the organisations to perform activities related to this. Data Integration and Management (DIM) capability is defined as "the ability of organisations to utilise tools and techniques to collect, integrate, transform and store data from heterogeneous data sources". The role of data integration capability is investigated in a few studies such as (Gupta and George 2016). When data is stored in silos it prevents the data users from getting access to vital information, and hence data integration is an important practice to achieve a status of 'single source of truth'. Moreover, advanced database technologies such as NoSQL and Hadoop are widely used to tackle data integration and storage problems. So, to reflect these aspects, the items used to measure this construct are adapted from Gupta and George (2016) and several other sources given in Table 4.4. Accordingly, respondents are asked to express their level of agreement or disagreement on the following statements regarding the DIM capability of their organisation: (a)We integrate data from multiple internal sources into a data warehouse or data mart for easy access, (b) We integrate external data with internal data to facilitate high-value analysis of our business environment, (c) We have the ability to Extract, Transform and Load (ETL) data from across systems and organisational boundaries, (d) In our organisation, data is not integrated or poorly integrated (Reverse coded), (e) Our data storage system is able to manage different data types beyond structured data, (f) Our data storage system is able to manage large volume of data, (h) We have adopted parallel computing approaches (e.g., Hadoop) for big data processing, (i) We have adopted new forms of databases such as Not Only SQL (NoSQL) for storing and retrieving of data.

4.8.1.3 Items for the advanced analytics construct

The advanced Analytics capability is defined as the ability of organisations to utilise tools and techniques to analyse data in batch wise, real-time, near-time, or as it flows and extracts meaningful insights for decision making. As discussed in the literature review chapter, there are different types of analytic such as descriptive, diagnostic, predictive and prescriptive, that can be used in the management of various business functions such as marketing, operations, sales, and customer relations, supply chain management etc. This construct aims to measure the organisation's general ability to use these analytics techniques. The items used to measure this construct are: (a) we use analytics to get data-driven insights into our historical business performance, (b) we use analytics to predict future events of our business environment, (c) we use analytics to prescribe possible courses of action for our business, (d) the performance of analytic models is regularly reviewed once deployed, (e) we are able to use analytical tools to convert data into actionable insights.

4.8.1.4 Items for the digital analytics construct

Digital analytics is defined as the ability of an organisation to analyse data generated from the digital environments such as websites and social media. The items used in this construct measure the ability of organisations to analyse clickstream data and information sourced from the web. The scales to measure this construct are fully adapted from (Cao et al. 2015), and the respondents are asked to indicate how often the following techniques are used in their organisation on the scale 1 (never) to 5 (always): (a) web analytics, (b) social media analytics, and (c) text, audio, video analytics.

4.8.1.5 Items for the data visualisation construct

The data Visualisation capability construct is defined as "the ability of organisations to utilise tools and techniques to render information visual and deliver the data-driven insights intuitively in a timely manner to the decision makers" (Arunachalam et al. 2016). The measures for this construct are adapted from several sources including (Popovič et al. 2012), (Radcliffe, 2014), (Sulaiman et al. 2015) and (Wang et al. 2016b). Accordingly, respondents are asked to indicate their level of agreement or disagreement with the following statements: (a) we have adopted data visualization tools, (b) data-driven insights are delivered using dashboards or other interactive visualisation tools, (c) data-driven insights are delivered in such a way that they are easily understandable by the target group, (d) data-driven insights are delivered in real-time, (e) data-driven insights are shared seamlessly across our organisation, regardless of the location. These items are intended to measure the existence of data visualisation capabilities such as interactive dashboards, mechanisms for the delivery of data-driven insights, and tools to deliver data-driven insights in real-time.

4.8.1.6 Items for the data-driven culture construct

Data-driven culture is an important component of BDA practice. It symbolises the intangible resource of an organisation representing the beliefs and opinion of employees and top management regarding data-driven decision making. Traditionally, decision-makers rely on heuristics based decision-making practices (Busenitz and Barney 1997), which are subject to biases whereas, data-driven decision-making allows decision makers to make objective decisions. However, it can be argued that the willingness to change and accept data-driven culture is imperative for the BDA practice. Hence, this study has adapted scales from Sangari and Razmi (2015) and Cao et al. (2015) to measure the presence of a data-driven culture within the organisation. Accordingly, respondents are asked to express their views on the following statements on the scale of 1-strongly disagree to 5-strongly agree: (a) we consider data as a tangible asset, (b) we are willing to override our own intuition when data contradicts our viewpoints, (c) we base our decisions on data rather than our instinct, (d) we continuously assess and improve our practices in response to insights extracted from data, (e) we continuously coach our employees to make decisions based on data.

4.8.1.7 Items for big data skills construct

Big Data Skills refers to the know-how of utilising big data technologies to extract data-driven insights (Gupta and George 2016). Davenport and Patil (2012) argued that while more emphasis is given to understanding how to use big data technologies such as ‘Hadoop’ and ‘MapReduce’ to extract insights, similar importance has to be given to developing people with Big data skill sets. Analytical skills include various topics such as statistics and database skills, data integration skills such as performing ETL operations and query processing (Chen et al. 2012; Schoenherr and Speier-Pero 2015). In this research, Big Data skills are measured on various aspects such as training and recruitment of skilled personnel and the organisation’s ability to possess skilled resources for performing big data analytics tasks. Accordingly, all the items for measuring this construct are adapted from (Gupta and George 2016), and the respondents are asked to indicate their view on the following statements on the scale of 1-strongly disagree to 5-strongly agree: (a) we provide big data analytics training to our own employees, (b) we hire new employees that already have big data analytics skills, (c) our big data analytics staff have the right skills to accomplish their jobs successfully, (d) our big data analytics staff have suitable education to fulfil their jobs, (e) our big data analytics staff holds suitable work experience to accomplish their jobs successfully, (f) Our big data analytics staff are well trained.

4.8.2 Operationalising the dimensions of ACAP

Absorptive Capacity (ACAP), is a multi-level multi-dimensional construct (Roberts et al. 2012), used in literature predominantly to explain the mechanism of organisational learning. So, ACAP can be used to measure the learning capability of any entities: an organisation or a team or even a country. In this research, the focus is on manufacturing firms' learning abilities and ACAP is measured at the organisational level. Moreover, the concept is multidimensional in nature and it is measured in various ways depending on the context, as discussed in the literature review (section 2.6). However, the focus of this research is related to technology, and similar to previous studies such as (Jean et al. 2008), (Hefu Liu et al. 2013b), (Kauppi et al. 2013) and (Wei et al. 2015) ACAP is conceptualised as a consequent of technology practice. Moreover, the scales for the ACAP construct is adapted entirely from (Liu et al. 2013), who conceptualised it as having four dimensions (acquisition, assimilation, transformation, exploitation) that are independent of each other. However, some scholars such as Zahra and George (2002) suggested that the four dimensions of ACAP can be grouped into two factors such as potential ACAP (PACAP) and Realised ACAP (RACAP). This research intends to measure all the four dimensions of the ACAP construct. Consequently, the organisation's ability to acquire, assimilate, transform and exploit knowledge obtained from BDA practice is measured using the scales adapted from (Liu et al, 2013b). First, *acquisition* denotes the ability of the organisation to acquire knowledge and measured using the following items: (a) we are successful in learning new things, (b) we are effective in developing new knowledge or insights that have the potential to influence product/service development, (c) we are able to identify and acquire internal (e.g., within the firm) and external (e.g. market) knowledge. Second, *assimilation* denotes the ability to absorb and comprehend the acquired knowledge and measured using the following items: (a) we have effective routines to identify, value, and import new information and knowledge from partners, (b) we have adequate routines to analyse the information and knowledge obtained, (c) we have adequate routines to assimilate new information and knowledge. Third, *transformation* denotes the ability to integrate the prior knowledge and new knowledge acquired and it is measured using the following items: (a) we can successfully integrate our existing knowledge with the new information and knowledge acquired, (b) we are effective in transforming existing information into new knowledge, (c) we can successfully grasp the opportunities for our firm from new external knowledge. Finally, *exploitation* denotes the ability to utilise and commercialise the knowledge generated to achieve firm objectives such as new product developments, and it is measured using the following items: (a) we can successfully exploit the new integrated information and knowledge into concrete applications, (b) we are effective in utilising knowledge into developing new products, (c) we constantly consider better ways to exploit knowledge.

Table 4.5 Scales for ACAP construct

Dimensions		No of items	Scale used	References
Potential ACAP (PACAP)	Acquisition	3	Likert scale	(Liu et al. 2013b)
	Assimilation	3		
Realised ACAP (RACAP)	Transformation	3	Likert scale	(Liu et al. 2013b)
	Exploitation	3		

4.8.3 Operationalising the dimensions of DIQ

Several researchers have claimed that data and information quality can be assessed using dimensions such as completeness, timeliness, reliability, consistency and accuracy (Pipino et al., 2002; Sidi et al., 2012; Wand and Wang, 1996). These dimensions reflect the ability to maintain data and information quality. The subject of this construct is the availability of quality data for decision-making to managers in the organisation. It is argued that quality data can help workers to identify any changes in processes and allows them to take corrective actions before defective products are produced (Kaynak 2003; Kim et al. 2012). In this research, the DIQ construct is measured using scales adapted from (Warth et al. 2011). The respondents are asked to indicate their ability to maintain the quality of data products/information from poor (1) to excellent (5) on the following aspects: (a) completeness, (b) timeliness, (c) reliability, (d) consistency, (e) accuracy.

4.8.4 Operationalising the dimensions of SCA capability construct

The construct Supply Chain Analytics (SCA) capability aims to measure the ability of organisations to use big data analytics on various activities of supply chain management such as sourcing, purchasing and forecasting. While, the BDA maturity construct measures the ability of an organisation to use it at an intra-organisational level, the SCA capability construct measures to what extent the organisation uses it to manage the supply chain functions. All the items of this construct are directly adopted from an existing study (Chen et al., 2015). Accordingly, respondents are asked to indicate to what extent BDA is used in the following supply chain functions: (a) sourcing analysis, (b) purchasing spend analysis, (c) CRM/customer analysis, (d) network design/optimisation, (e) warehouse operations improvements, (f) process monitoring, (g) production process optimisation, (h) logistics process improvements, i) forecasting/demand management, and (j) inventory optimisation.

4.8.5 Operationalising the dimensions of operational performance

Operational performance is related to the performance of the operations function, rather than the performance of the whole business (Liu et al. 2012). The term 'operations management' refers to "all of the activities, decisions and responsibilities of managing the production and delivery of products and services" (Boddy 2014, p.566). This research intends to use 5 different variables to measure operational performance. According to literature the most commonly used variables to measure operational performance are product quality, cost, flexibility, time and delivery (Gunasekaran et al. 2001; Neely et al. 1995; Shepherd and Gunter, 2006; Beamon, 1999). In the context of analytics capabilities, Chae, Yang, Olson, et al. (2014) measured its impact on operational performance based on the flexibility and delivery dimensions. Jamehshooran et al. (2015a) and Jamehshooran et al. (2015b) have tested the influence of analytics capabilities on delivery, cost, quality, reliability and flexibility dimensions of operational performance. However, there is no literature that has investigated the impact of BDA capabilities from a maturity perspective on operational performance in the context of the UK manufacturing sector. In this research, the operational performance dimensions are measured using scales adapted from two studies (Liu et al. 2012) and (Wu and Chuang, 2010).

Table 4.6 Scales for operational and innovation performance constructs (dependent variables)

Dimensions		No of items	References
Innovation performance		3	(Wang et al. 2011) (Chen et al. 2015b)
Operational performance	Product Quality	3	(Wu and Chuang, 2010) (Liu et al. 2012) (Peng et al. 2008)
	Cost	3	(Liu et al. 2012) (Peng et al. 2008)
	Flexibility	3	(Liu et al. 2012)(Peng et al. 2008)
	Time	4	(Liu et al. 2012)(Peng et al. 2008)

Product Quality

The quality of product and services is determined by the requirement of the customers (Boddy 2014). Therefore, any product or service that conforms to the requirement of the customers can be considered as a quality product. Moreover, the functionality and durability of the product can also be considered as important dimensions of quality. Hence, the following scales are adapted from Liu et al. (2012) and Wu and Chuang (2010) as they reflect the subject of the construct. The respondents are asked about their organisation's performance on the following parameters on the scale of 1-poor to 5-excellent: (a) product quality and performance, (b) conformance to product specifications (meet established standards/customers' requirements), (c) products reliability (probability of a product malfunctioning/failing within a specified time).

Cost

As discussed in the literature review, managing cost is one of the elements that determine the competitiveness of the manufacturing firms. Cost management is the central focus of organisations (Blocher et al. 2010), and they can adopt various strategies to manage cost such as cost leadership strategy or cost differentiation strategy (Porter 1985). Due to global competition, firms strive to produce high-quality products at low cost, and its dimensions can include transportation costs, rental costs, etc. In this research, the following dimensions of cost adopted from Liu et al. (2012) are used: (a) unit cost of manufacturing, (b) inventory costs, (c) overhead costs, (d) price competitiveness. These dimensions are intended to measure how effective and competitive the firms are in terms of managing cost.

Flexibility

Scholars have recognised manufacturing flexibility as a critical component of competitive advantage (Vokurka and O'Leary-Kelly 2000a), and it is often associated with the changing environment. Flexibility is perceived as a multidimensional construct (Oke 2013b; Sethi and Sethi 1990; Vokurka and O'Leary-Kelly 2000b). Wei et al. (2017) argued that flexibility denotes a firm's ability to reorganise internal resources to reduce constraints, mainly related to labour flexibility (P C Patel et al. 2012), changing the design and delivery necessities (Vokurka and O'Leary-Kelly 2000b). In this research, flexibility is measured with a focus on the product, rather than other dimensions such as labour, delivery, etc. Consequently, the following scales are adapted from Liu et al. (2012) and Peng et al. (2008): (a) flexibility to change product mix, (b) flexibility to change volume, (c) ability to produce customized product features. These scales are intended to measure the ability of the firm to customise product features and volume coping with the changing market demands.

Time

Time-based operational performance is argued to be externally focused (Koufterosa et al. 1998). It measures the effectiveness of operational systems in reducing response time and fulfilling customer demands. It measures various aspects of operational systems such as responsiveness to demand, adjustment of manufacturing lead-time and quick delivery. To compete on time, organisations have to be agile and redesign processes, and the speed at which the organisations react to the changing demands determine their competitiveness. So, based on Liu et al. (2012) and Peng et al. (2008), the time-based performance dimension is measured using the following items on the scale of 1-poor to 5-excellent: (a) order fulfilment lead-time, (b) manufacturing lead-time, (c) supply chain throughput time, and (d) on-time delivery performance.

4.8.6 Operationalising the dimensions of innovation performance

The field of innovation research is extensive (Damanpour 1991). Previous studies have measured innovation performance based on the ratio of annual sales originating from new products and service (Kostopoulos et al. 2011). Although this approach can provide a more objective assessment of innovation, it only captures the product/service innovation and there is a lack of accessibility to this information. Moreover, in operations management research, innovation is often conceptualised as management/administrative innovation, technology innovation or product/service innovation (Kim et al. 2012; Mazzola et al. 2015; Wang et al. 2011). The impact of BDA on innovation performance is not empirically investigated in the previous literature. Accordingly, respondents are asked to indicate their level of agreement with the following items: (a) we achieve substantial innovations in our product and/or service offerings, (b) we achieve substantial innovations in our management practices, and (c) we achieve substantial innovations through adopting manufacturing technology. Using the scales above, an organisation’s complete innovation level is measured by capturing all three aspects of innovation.

4.8.7 Control variables

Control variables are the “observable and measurable variables that need to be kept constant to avoid them influencing the effect of the independent variable on the dependent variable” (Saunders et al. 2016, p.179). In this study, two control variables are used in the conceptual framework: namely the number of employees and the annual turnover. Both these variables can be used to determine the firm size and can influence the BDA capabilities and operational and innovation performance. These are some of the most commonly used firm size variables in operations management literature (Cai et al. 2016; Sánchez and Pérez, 2005). It can be argued that firm size may have a significant effect on BDA practice because large firms benefit from the availability of resources whereas small firms can quickly adopt new technology. Thus, to assess the influence of these control variables, respondents are asked to specify their number of employees and the annual turnover of their organisation.

4.9 Questionnaire instrumentation and design

Once the scales to measure all the constructs in the model are identified from the literature, the questionnaire design process is carried out. A good survey instrument should contain valid and reliable measures that have a proper layout and are worded without confusion to attain an increased response rate (Millar and Dillman 2011; Dillman et al. 2014; Forza 2002). Neuman (2014) has provided a list of features to avoid ensuring clarity and prevent response bias. Accordingly, in this research, the survey instrument is prepared avoiding abbreviations, double-barrelled questions,

questions with more than 20 words and jargon. Moreover, as suggested by Forza (2002), to eliminate the tendency for participants reflexively choosing options on the questionnaire, reverse worded questions are included in the questionnaire. However, reverse coding of these reverse worded questions is carefully performed before the commencement of statistical analysis (Bryman and Bell 2011).

Apart from the wording, the survey layout is crucial for an online survey as there is no interaction between the researcher and the respondents (Sue and Ritter 2007). So as to ensure the survey layout is appropriate, respondents are provided with instructions for completing the survey. The online survey is designed using 'Qualtrics', an online platform to design and distribute questionnaire surveys. Qualtrics allows the researcher to control the number of choices the respondents should choose. Further, for statistical purposes, most of the questions are measured on a 5-point Likert scale. To ensure flexibility, the questionnaire is designed such that participants can move back to previous pages to alter their responses. To avoid fatigue from excessive scrolling through (Sue and Ritter 2007), a multipage design of the questionnaire is used in this research. This approach may avoid premature termination and subsequently reduce the percentage of incomplete surveys. Furthermore, it is made sure that the questionnaire is consistent in terms of font size, text, background style and colour (Forza 2002).

The survey contains several sections: (a) introduction and definition of Big Data Analytics; (b) background information (Industry type and screening questions); (c) Big Data Analytics capabilities; (d) Organisational learning capability (Absorptive capacity construct); (e) Firm performance; (f) demographic information; (g) thank you and final notes. A copy of the questionnaire is provided in the Appendix A. In the first section, a cover letter is attached, and a jargon-free statement of Big data analytics is provided as industry professionals refer to this concept using different names such as 'Business analytics' and 'Business Intelligence'. Predominantly, the question types used in the survey are closed-ended in nature that allowed the researcher to collect responses on a fixed set of categories. Advantages of using this closed-ended approach include ease of coding, ease of comparing responses of different informants and prompt response (Neuman 2014; Forza 2002). However, the possible drawback of closed-ended questions is the limited choice of answers respondents can choose from. Furthermore, the majority of questions included in the survey are on the 5-point Likert scale, and the demographic questions are appropriately designed to include all the possible choice that the respondents can choose from.

4.10 Questionnaire validity and reliability

Reliability and validity are two important issues of concern in research that utilises the survey instrument to measure constructs (Meredith et al. 1989). "Reliability is the extent to which a test or procedure produces similar results under constant conditions on all occasions" (Bell and Waters 2014, p.121). Reliability indicates consistency, accuracy, stability (Kerlinger 1986), and also denotes inter-correlation between the items that measure the same construct (Saunders et al. 2016). De Vaus (2002) argued that questions used to measure a construct can be unreliable due to poor wording, or questions not being specific to respondents' contextual factors such as educational background and ethnicity. There are several ways to assess reliability: 1) test-retest methods, 2) alternative forms method, 3) Cronbach's alpha coefficient or 4) Werts, Linn, and Joreskog (WLJ) composite reliability method (O'Leary-Kelly and J. Vokurka 1998). The first two methods (test-retest and alternative forms) are similar in approach, they involve measuring a construct at different time periods 't' and 't+1', but are often criticised due to their theoretical and practical drawbacks (O'Leary-Kelly and J. Vokurka 1998). However, Cronbach's alpha and WLJ composite reliability method are widely used to ensure reliability. In this research, to ensure reliability, the questions that contain poor wording or phrases are removed and the multi-item measurement method recommended by De Vaus (2002) is used. Moreover, the reliability of the constructs is also tested using Cronbach's alpha coefficient (Cronbach and Meehl 1955) during Exploratory Factory Analysis (EFA) and by the WLJ method. The WLJ method uses Confirmatory Factor Analysis (CFA) to drive composite reliability scores. Both Cronbach's alpha and the composite reliability index can range from 0 to 1, with a value closer to 1 indicating high reliability. However, Nunnally (1978) has mentioned that the threshold criteria of Cronbach's alpha coefficient can vary depending on the purpose of the study. Accordingly, an instrument used in management research should achieve reliability of at least 0.7, with a lowest acceptable level of 0.6 (Nunnally 1978). However, in the case of applied research, when the outcome of the research would affect an individual's existence, a reliability of above 0.90 or even 0.95 is desirable (Nunnally 1978). Thus, in this study, a reliability value above 0.70 is considered as acceptable.

On the other hand, validity denotes to the extent to which a concept or a phenomenon is accurately measured. Validity refers to "whether a measure accomplishes its claims" (Cooper and Schindler 2014, p.201) or "the scientific utility of a measuring instrument" (Nunnally and Bernstein 1994, p.83). The two major forms of validity are internal and external validity. Zikmund et al. (2010, p. 274) stated that "Internal validity exists to the extent that an experimental variable is truly responsible for any variance in the dependent variable". Internal validity concerns about causality. However, external validity refers to the generalisability of the findings. Selecting a sample that truly represents the target population will improve external validity. In this research, the data sample represents the UK manufacturing companies and the findings can be generalisable to this context. Moreover, there are various other forms of validity discussed in the literature (Table 4.7).

Table 4.7 Various forms of validity and measurement methods

Types	What Is Measured Degree	Methods
Content	The degree to which the content of the items adequately represents the universe of all relevant items under study	• Judgmental • Panel evaluation with content validity ratio
Criterion-Related	The degree to which the predictor is adequate in capturing the relevant aspects of the criterion.	• Correlation
Concurrent	Description of the present; criterion data are available at the same time as predictor scores.	• Correlation
Predictive	Prediction of the future; criterion data are measured after the passage of time.	• Correlation
Construct	Answers the question, "What accounts for the variance in the measure?"; attempts to identify the underlying construct(s) being measured and determine how well the test represents it (them)	• Judgmental • Correlation of the proposed test with established one • Convergent-discriminant techniques • Factor analysis (EFA and CFA) • Multitrait-multimethod analysis

Source: (Cooper and Schindler 2014)

Content validity

Content validity refers to "the extent to which the indicators measure the different aspects of the concept" (De Vaus 2002, p.54). Content validity, also known as 'face validity' (Hair et al. 2010), is considered as good "if the instrument contains a representative sample of the universe of the subject matter of interest" (Cooper and Schindler 2014). There are two ways to validate the content of a questionnaire: 1) determine the validity of the content based on judgement, 2) use an expert panel to assess the validity of the content. Moreover, an extensive literature review can also be used to ensure content validity. Accordingly, in this research, the validity of the content is assessed by using an extensive literature review, self-judgement, and with experts from both academic and industry**.**

Predictive / concurrent / criterion validity

These types of validity refer to the relationship between the scores of predictor and criterion variables. Regarding predictive validity, Nunnally(1978) stated that "When an instrument is intended to perform a prediction function, validity depends entirely on how well the instrument correlates with what it is intended to predict (a criterion)" (Nunnally 1978). As mentioned in Table 4.7, correlation and canonical correlation are the two important techniques used to assess these types of validity (Cooper and Schindler 2014; Flynn 1990). Consequently, in this research, squared multiple correlations (R-square), which is the square of the correlation coefficient between predictor and criterion, are obtained to assess the predictive validity.

Construct Validity

Schwab (1980) noted that the definition of a construct is the crucial and first step in the validating process of a construct. The author argued that in the process of validation, the researcher should have clarity on the multidimensionality nature and number of dimensions of the construct. (O'Leary-Kelly and Vokurka 1998). Bagozzi and Yi (2012, p. 19) stated that "Construct validity is the extent to which indicators of a construct measure what they are purported to measure. Unlike reliability, which is limited to the degree of agreement among a set of measures of a single construct, construct validity addresses both the degree of agreement of indicators hypothesized to measure a construct and the distinction between those indicators and indicators of a different construct(s)." Accordingly, in this research, construct validity is measured using goodness-of-fit indices. Moreover, the convergent and discriminant validity of the constructs used in this research are also assessed.

4.11 Quantitative data collection

4.11.1 Population and sampling strategy

4.11.1.1 Unit of Analysis

Understanding the level of analysis is important for increasing the focus on the research problem and recognising the types of data to be collected (Flynn 1990). The unit of analysis refers to "the level of aggregation of the data collected during the subsequent data analysis stage" (Sekaran 2003, p.132). Similarly, a unit of the analysis indicates "what or who is [being] studied" (Babbie, 1992, p. 92 as cited in Rungtusanatham et al. 2003). Here, the unit of analysis is UK manufacturing organisations, and the individual responses collected from participants indicate the capabilities of the organisations. The unit of analysis, data collection, and interpretation of findings have to be consistent as it is critical for theory development and the generalisation of the findings (Saunders et al. 2009).

4.11.1.2 Selection of potential respondents

Identification of potential respondents is substantial as it determines the quality of the responses. A respondent should be someone who has knowledge about the phenomenon of the study and is willing to share information. The focus of this research is on investigating a firm's big data analytics and other capabilities such as absorptive capacity. Senior executives of the UK manufacturing companies are identified as potential respondents of this study as they are anticipated to possess both the knowledge of their firms' capabilities as well as performance. Moreover, senior executives are mostly accountable to make un-programmed decisions such as investment in BDA technology (Boddy 2014). So, senior executives such as CEOs, Chief Operations officers, IT managers, SC managers are selected as key respondents.

4.11.1.3 Sampling strategy

After identifying the unit of analysis, target population and the potential participants, the next stage concerns selecting the sampling strategy. Surveying the entire population to answer the research questions is impractical and constrained with time and budget, which emphasises the necessity of sampling the population (Saunders et al. 2016). There are several ways of sampling discussed in the academic literature, but broadly speaking, there are only two major types: probability or representative sampling, and non-probability sampling. Rungtusanatham et al. (2003) found that the majority of scholars in operations management have utilised probability sampling. Simple random sampling, systematic sampling, stratified sampling and clustered sampling are the techniques that belong to probability sampling (Cowles and Nelson 2015).

According to Dillman et al. (2014), in simple random sampling, an equal chance of inclusion is given to all the members of the population. Members are randomly selected, and the selection is independent. In systematic random sampling, the first element is selected randomly, and every n^{th} number is selected systematically. Whereas, in stratified sampling, the population is divided into strata (groups) before the random selection. Stratified random sampling provides control over representativeness of specific groups of interest. Cluster sampling involves grouping or clustering sample units and then randomly selecting samples from a chosen cluster. The main difference between stratified and cluster sampling is, in stratified sampling, strata (groups) are created and a random sample is drawn from each group. However, in the case of cluster sampling, the sample is grouped/clustered and then specific clusters are used to sample data, e.g. based on geographical areas and manufacturing firms. The advantages and disadvantages of all the probability sampling techniques are given in Table 4.8.

Table 4.8 Advantages and disadvantages of probability sampling techniques

Sampling design	**Advantages**	**Disadvantages**
Simple random sampling	Generalisability of the findings is high with this sampling. Limited knowledge of the population is sufficient to perform simple random sampling. Computing error and analysing data can be done with ease.	Not as efficient as stratified sampling. Compared to stratified sampling, the possibility of more errors. Respondents may be widely dispersed, hence cost may be higher
Systematic sampling	Easy to do sampling when population frame is available. Moderate cost compared to other techniques.	Systematic biases is a disadvantage
Stratified random sampling	Most efficient among all probability designs. All groups are adequately sampled and comparisons among groups are possible	Stratification must be meaningful. More time-consuming than simple random sampling or systematic sampling.

		Population frame for each stratum is essential. if stratified lists are not already available, they can be costly to prepare.
Cluster sampling	In geographic clusters, costs of data collection are low	The least reliable and efficient among all probability sampling designs since subsets of clusters are more homogeneous than heterogeneous. The researcher must be able to assign population members to unique cluster or else duplication or omission of individuals will result

Sources: (Sekaran 2003), (Cowles and Nelson 2015) and (Zikmund et al. 2010)

The Fame database is used to identify and gather email contacts of the potential respondents. By utilising Standard Industrial Classification **(**SIC) codes, only the firms that are classified under manufacturing category are considered in this study. These categories include companies that produce products ranging from machinery, electronic equipment, chemical products, food, automobile, etc. The choice is purposefully made with an assumption that these organisations operate in a dynamic competitive environment and would possess some form of BDA capabilities.

Moreover, the focus of this study is manufacturing companies in the UK. Fowler Jr, (2014) mentioned five critical issues that must be considered before sampling: 1. a decision on whether to use probability sample or not, 2. the sample frame, 3. the sample size, 4. the sample design, 5. the response rate. The population for the quantitative survey can be selected through systematic random sampling. However, a sampling frame is a must to perform systematic probability random sampling. A sampling frame refers to “a complete list of all the cases in the target population from which your sample will be drawn” (Saunders et al. 2016, p.277). The researcher has used the ‘Fame database’, which is widely used in academic researches that focus on the context of the UK. Although the Fame database does claim to have email contacts of all the companies in the UK, it is unauthenticated. Hence, in this study, a simple random sampling approach is used to identify the target sample.

4.11.2. Questionnaire administration and data collection

4.11.2.1 Pilot study

After the methodical design of the survey process, the measurement instrument is pre-tested to highlight problems before the initiation of the main data collection (Malhotra and Grover 1998). Pretesting of the survey instrument is an integral part of the survey design which provides insights into the layout and wording issues that decreases the ease of completion (Flynn 1990). In this research, the pilot study is conducted using the procedure proposed by (Forza 2002). The procedure suggests testing the survey instrument using three groups of people such as academics, field experts and potential informants (Figure 4.9). The objective of the pilot test is to verify whether the questionnaire is in accordance with the study's objective. It is suggested that random sampling is not mandatory for a pilot study (Flynn 1990), and hence convenience sampling is used to identify participants for the pilot study. As shown in Figure 4.9, a pilot study is conducted with both academic and industry experts.

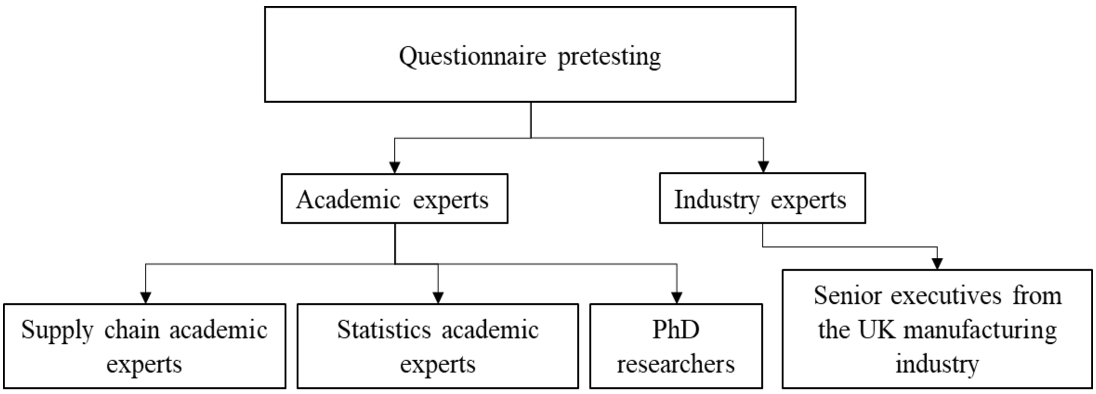


Figure 4.9 Pretesting of the survey instrument

In the first stage, 5 supply chain academic experts within the Operations Management and Decision Science (OMDS) division, 3 statistics experts from Sheffield Methods Institute (SMI) and Sheffield Management school, and 5 PhD researchers are invited to pilot the test. As discussed in section 4.8, the survey instrument is developed in the 'Qualtrics' online platform. A link of the draft online questionnaire is sent to the academic experts and a face-face semi-structured interview is arranged to discuss and collect feedback on the questionnaire. The questions asked in the semi-structured interview are provided in Appendix C. The purpose of conducting a pilot test with academic experts is to examine the face validity of the measures especially in the case of the BDA constructs (Hair et al. 2010; Dillman et al. 2014). Both wordings and layout of the survey are examined, and the content is validated with the feedback from the academic experts in the field of supply chain technology and operations management.

In the second stage, the questionnaire is revised as per the suggestions of the academic experts. For instance, based on the feedback, the length and format of the questionnaire is revised and a brief description of Big Data is added. Then, the online link to the revised questionnaire is sent to 25 senior executives of manufacturing companies located in the Yorkshire region. The email contacts of these 25 industry experts are acquired from the Sheffield University Management School. The semi-structured interview questions are also included at the end of the questionnaire to allow the industry experts to provide feedback. The main objective in this stage of the pilot test is to verify that the wordings of the questionnaire is completely jargon-free and understandable by the industry practitioners and ensure that the potential informants are comfortable with the length of the questionnaire as well. Participants are asked to provide feedback on any aspect of the questionnaire that is unclear and hard to understand. Finally, 14 industry experts responded to the pilot survey. Based on their feedback, a few final changes are made, and it is noticed that it took an average of 15-16 minutes to complete the questionnaire.

4.11.2.2 Main survey

This section is concerned with the procedures used for administering the main survey after the pilot test. The rationale for choosing a web-based survey administration is discussed in section 4.7. In addition, Sue and Ritter (2007) highlighted that there are three options for survey distribution via email: (1) send the questionnaire as an attachment to an email, (2) include the questionnaire in an email message, (3) embed an online survey link with the email invitation. The first option is constrained due to the fear of malware threats, and the second option is only suitable for a short survey. So, the third option is adopted in this study. Using this method, the anonymity of the participant is ensured.

Consequently, potential participants are contacted via an invitation email. Based on the perspective of social exchange theory, Sue and Ritter (2007) argued that the respondent's decision to participate in the questionnaire survey is primarily based on the ability of a researcher in convincing and motivating them. The invitation email indicated the collaboration of the University of Sheffield and confirmed that the motive of the study is purely academic, not commercial. The invitation email is signed by two senior lecturers of the Management School as that might positively influence the response rate. The invitation email explained the aim of this research and the benefits it would bring to the UK manufacturing sector. Also, to eliminate response bias and increase the response rate, the invitation email addresses the confidentiality of the responses. Moreover, Dillman et al. (2014) suggested some strategies for effective administration of a survey: (a) create a questionnaire that is interesting to read, (b) eliminate any necessity of mental or physical efforts to complete the questionnaire, (c) recompense the respondent either via tangible awards or positive regards, (d) remove any elements of monetary cost to participants. This research follows the guidelines of Dillman et al. (2014), and the respondents are informed of the prize draw to win £100 in appreciation of their participation. Achieving a high response rate via an online survey method can be challenging (Sue and Ritter 2007). So, follow up emails and telephone calls are made to enhance the response rate. Accordingly, the first email invitation is sent on the beginning of April 2017 and the final reminder is sent on the end of October 2017. A detailed information on the events occurred during the data collection stage is given in Chapter 5 (section 5.2.1).

4.11.2.3 Ethical consideration

The ethical practice of research is a high priority in this study, as it is an important consideration in social science research (Saunders et al. 2009; Rea and Parker 2014; Creswell 2014). This study collects critical information about the capabilities and performance of organisations. Such information must be dealt with confidentiality as it is a source of competitive advantage for the participating firms. In this research, the data collection is conducted complying with the code of ethics of research practice. The researcher gained ethical approval from the ethics committee of the University of Sheffield before the initiation of the pilot study and main data collection. The ethical approval letter from the University of Sheffield is attached in Appendix B. Moreover, in the invitation letter, the respondents are assured of anonymity and security of the information that they provide via participating in the survey.

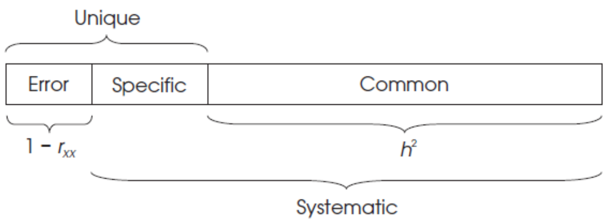
4.12 Quantitative data analysis

The quantitative data collected through questionnaire survey is analysed using two different techniques: 1) Structural Equation Modelling (SEM), and 2) Cluster analysis. The first-order constructs that are used to specify the second-order BDA

Maturity construct are subjected to internal consistency and reliability tests using Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis(CFA). SEM is widely used to explain the causal relationship between the second-ordered constructs in the conceptual model, and it is also used by researchers in the domain of BDA (Cao et al. 2015). On the other hand, clustering is a data mining technique used to identify homogeneous groups within the sample population. Consequently, the phenomenon of the digital divide between large organisations and SMEs is investigated using clustering techniques such as K-medoids, Hierarchical, latent class analysis and Model-based clustering algorithms. Cluster analysis is used to demonstrate the presence of the digital divide by grouping organisations into categories based on their BDA Maturity and profiling them with their firm attributes such as firm size and annual turnover. Cluster analysis techniques are widely used in academic literature to address similar issues (Lukman et al. 2011; Malhotra et al. 2005). The software package used to perform EFA is IBM SPSS, for the CFA and SEM analyses SPSS AMOS is used, and R (open source) is used for cluster analysis.

4.12.1 Factor analysis

In this study, some of the measures related to BDA capabilities and operational performance dimensions are adapted from multiple sources and their factor structure is unknown. So, factor analysis is used to address this issue. The main purpose of factor analysis is to identify "the underlying structure among the variables in the analysis" (Hair et al. 1998, p.92). Factor analysis is concerned with explaining the total 'common variance' between the variables and the factors (Mulaik 2010). There are two types of factor analysis: Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA). In both factor-analytic methods, the standardised indicator variance is partitioned as shown in Figure 4.10. The total variance is the sum of the unique variance and the systematic or true variance. The main purpose of EFA is "to explore the field, to discover the main constructs or dimensions" (Kline 1994). However, in CFA, factors have to be specified based on prior knowledge from existing studies or theory. So, in CFA, the researcher must specify: a) a number of factors, b) variables that reflect specified factors, c) whether the specified factors are correlated or not (Thompson 2004). Moreover, unlike EFA, CFA allows the researcher to test the fit of factor models. The graphical representation of the EFA model and CFA model or measurement model is given in Figure 4.11



Source: (Kline 2015)

Figure 4.10 Partition of standardized indicator variance in factor analysis

In this research, both EFA and CFA techniques are used. First, EFA allowed the researcher to explore the underlying factor structure in the data and identify factors and variables that correlate to form it. Second, CFA lets the researcher confirm the factor structure by validating it via various model fit measures. EFA and CFA are performed using IBM SPSS and AMOS software, respectively.

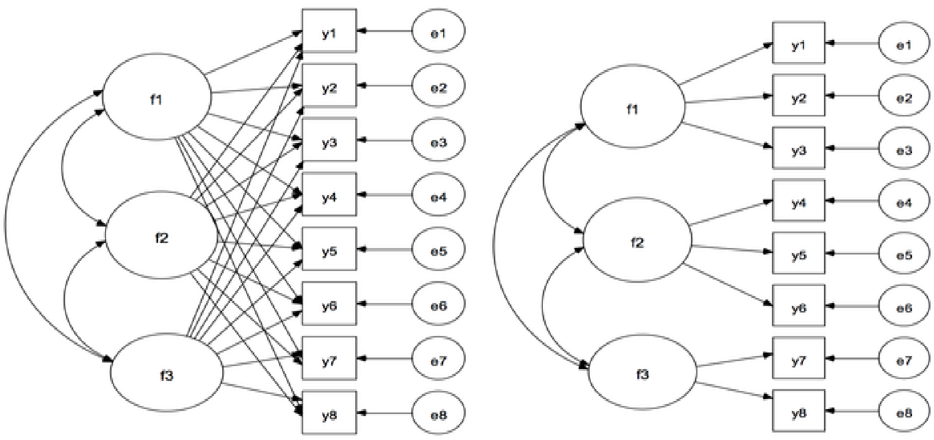


Figure 4.11 Three-factor EFA model (left) and CFA model (right)

4.12.1.1 Exploratory factor analysis.

Ever since its introduction by Charles Spearman, EFA has been commonly used in the academic literature mainly to find the interrelationship between the observed variables and group a set of variables that are highly correlated to form a factor (Loehlin and Beaujean 2017; Pituch and Stevens 2016; Kline 2015). Moreover, real data sets seldom produce a correlation pattern that is easy to distinguish, but "factor analysis provides the tools for analysing the structure of the interrelationships (correlations) among a large number of variables (e.g., test scores, test items, questionnaire responses) by defining sets of variables that are highly interrelated, known as factors" (Hair et al. 1998, p.92). EFA is a preliminary step to identifying latent variables that need to be studied using structural modelling (Loehlin and Beaujean 2017). So, in this research, EFA is conducted first before CFA.

Moreover, Fabrigar et al. (1999) argued that for implementing EFA, the researcher should decide on the following aspects: 1. what variables to be included in the study, 2. the number of variables to be included in the study, 3. the size and nature of the sample, 4. specific procedures such as factor extraction and rotation methods, 5. how many factors to expect. The decision on each of these criteria will have a significant impact on the outcome of factor analysis. Deciding on a factor extraction and rotation method is highly significant in EFA. There are various extraction techniques in EFA. However, principal component analysis, principal axes factoring and maximum likelihood are the most commonly used in the literature. The summary of extraction techniques, its goal of analysis and special features are provided in Table 4.9. Similarly, there are several types of rotation methods such as orthogonal and oblique rotation (Kline, 1994; Mulaik, 2010; Thompson, 2004; Yong and Pearce, 2013). Principal component analysis (PCA) with varimax rotation is widely adopted in both supply chain (Ağan et al. 2014; Gunday et al. 2011; Huo et al. 2014; Patel et al. 2012) and IT (Wang et al. 2015; Gupta and George 2016; Cao et al. 2015) literature to identify factors. However, Costello and Osborne (2005) argued that PCA is not a true factor analysis method and statistical theorists have a disagreement over the application of PCA. In fact, researchers such as Ford et al. (1986), Bentler and Kano (1990) and Floyd and Widaman (1995) have recommended not to use PCA over true factor analysis such as principal axes factoring and maximum likelihood. However, according to Thompson (2004, p. 50) “principal axes factor analysis is the same as principal components analysis, except that in principal axes factoring only the diagonal of the correlation matrix is modified by removing the 1.0's originally always residing there and iteratively estimating the true communality coefficients”. de Winter and Dodou (2012) argued that there is no evidence to claim that any ‘one’ extraction method is significantly better than another. However, the authors compared the usefulness of principal axis factoring (PAF) and maximum likelihood factor analysis (MLFA) using a simulated data and concluded that “MLFA is the most flexible method and best able to cope with severe model misspecifications” (p.708), and outperforms PAF in the case of a complex model. Fabrigar et al. (1999, p. 277) supported that use of MLFA and stated that “it allows for the computation of a wide range of indexes of the goodness of fit of the model [and] permits statistical significance testing of factor loadings and correlations among factors and the computation of confidence intervals.” Costello and Osborne (2005) suggested choosing PAF if the data severely violates the normality assumption, but if the data is relatively normally distributed then MLFA is the best choice. Hence, in this study, MLFA is used to extract factors.

Table 4.9 Purpose and special features of factor extraction methods

Extraction Technique	Goal of Analysis	Special Features
Principal components	Maximise variance extracted by orthogonal components	Mathematically determined, an empirical solution with common, unique, and error variance mixed into components
Principal factors	Maximise variance extracted by orthogonal factors	Estimates communalities to attempt to eliminate unique and error variance from variables
Image factoring	Provides an empirical factor analysis	Uses variances based on multiple regression of a variable with all other variables as communalities to generate a mathematically determined solution with error variance and unique variance eliminated
Maximum likelihood factoring	Estimate factor loadings for a population that maximize the likelihood of sampling the observed correlation matrix	Has significance test for factors; especially useful for confirmatory factor analysis
Alpha factoring	Maximise the generalizability of orthogonal factors	
Unweighted least squares	Minimise squared residual correlations	
Generalized least squares	Weights variables by shared variance before minimising squared residual correlations	

Source: (Tabachnick and Fidell 2012)

Subsequently, after the factor extraction using MLFA, the actual factor loadings has to be transformed using factor rotation methods (to form a rotated factor loadings). Factor rotation maximises the high correlations and minimises the low correlations between the variables and the factors (Tabachnick and Fidell 2012). It allows the patterns to be represented as a simple structure that is easy to interpret

(Bartholomew et al. 2011). The two broad groups of factor rotation methods are; orthogonal and oblique rotations. Orthogonal rotation is further classified into 'varimax', 'quartimax' and 'equamax'. Whereas, 'promax', 'direct oblimin' and 'quartimin' are the types of oblique rotations. The orthogonal rotation produces factors that are uncorrelated. Whereas, oblique rotation allows factors to be correlated (Tabachnick and Fidell 2012). Costello and Osborne (2005) claimed that oblique rotation produces more accurate results, as orthogonal rotation may result in the loss of valuable information. In this study, the factors in the conceptual model are not viewed as completely independent of each and hence oblique rotation (promax) is used (Fabrigar et al. 1999).

Sample size also plays an important role in EFA, as the correlation coefficients obtained from a small sample size are considered less reliable (Tabachnick and Fidell 2012). However, there are conflicting views regarding the sample size in management research. MacCallum et al. (1999) found out that a sample size of 100-200 is sufficient if most factors are well-defined in terms of factor loadings and communalities. The authors argued that when communalities, which are squared multiple correlations among the variables, are above 0.5 and factor loadings are above 0.8, 100-200 samples are sufficient, irrespective of the number of variables. However, if the communalities are less than 0.5 with poor factor loadings, a sample size of more than 500 is required. Moreover, variables with low communality will tend to load onto multiple factors and it is suggested they are removed from further analysis. As suggested, variables which have less than 0.5 communalities are removed from the analysis (De Vaus 2002). Costello and Osborne (2005) have suggested various possibilities of cases to items ratios such as 2:1, 5:1, 10:1, and 20:1. Nevertheless, the rules regarding sample size are disappearing (Costello and Osborne 2005). At the same time, the Kaiser-Meyer-Olkin (KMO) test (Kaiser and Rice 1974) and Bartlett's test of sphericity (Bartlett 1954) are widely used to measure the suitability of data for factor analysis. A value greater than 0.5 in the KMO test indicates that the sample is adequate for factor analysis (Field 2013). In more detail, Hutcheson and Sofroniou (1999) gave categories of acceptable KMO values; 0.5 - 0.7 are mediocre 0.7 - 0.8 are good, 0.8 - 0.9 are great, and value 0.9 and above are excellent. In the case of Barlett's test, which measures the multivariate normality, a significance value less than 0.05 indicates that the data is circa multivariate normal and acceptable for factor analysis (Pallant 2007;

Field 2013). So, in this research, the KMO test and Barlett's test are conducted to assess the adequacy of the sample before proceeding with factor analysis.

Deciding on the number of factors to be retained is also an important consideration in this research. Eigenvalues and scree plots are the most commonly used for this purpose. Factors with an Eigenvalue greater than 1.0 are retained in this research as suggested by Tabachnick and Fidell (2012). The Scree test is a simple graphical method to determine the number of factors (Costello and Osborne 2005). It involves the examination of the scree plot and looking for a natural bend on the elbow plot where the curve flattens out - the points above the elbow give the number of factors hidden in the data. In terms of factor loading, De Vaus (2002) suggested that, as a rule of thumb, it is unusual to use variables that have achieved a loading coefficient of less than 0.3. Several studies such as Gunasekaran et al. (2017) and Wang et al. (2015) that have used EFA, have considered the above-mentioned threshold level as a loading criterion. In this study, the researcher has made sure that the variables with communalities less than 0.5 and with poor loading of less than 0.5 are removed from the analysis. Finally, at the stage of EFA, the reliability of factors extracted is assessed using Cronbach's alpha coefficient (Nunnally and Bernstein 1994).

4.12.1.2 Confirmatory factor analysis

After reducing the observed variables into fewer factors or latent variables using EFA, Confirmatory Factor Analysis (CFA) has been used to verify and confirm the hypothesised model. According to Bagozzi et al. (1991), the CFA model is a powerful method for addressing construct validity as it makes fewer assumptions and provides more diagnostic information about reliability and validity.

Model specification is a significant step in CFA. If the measurement model is misspecified, then it will result in a poor model fit (Kline 2015). In this study, the measurement model is developed and the researcher specified a priori the number of factors, the relationship between the factors and indicators and the error correlations (Kline 2015). Moreover, models are specified in such a way the factors cause the variables or indicators, which is generally called a reflective measurement model. The inherent assumptions in CFA models are that both specific and error variance are contained in the error term.

As discussed in Chapter 3, this research aims to test hypotheses that evaluate the relationship between individual dimensions of BDA capabilities such as the data generation capability and the advanced analytics capability on the firm performance dimensions. In addition, the relationship between the overall BDA maturity (composed of 7 BDA capabilities dimensions) and the firm performance dimensions is evaluated. So, two different measurement models are developed and verified in this study. In the first measurement model, BDA capabilities are conceptualised as first-order constructs along with the mediating (ACAP, DIQ, SCA capability) and dependent variables (product quality, cost, flexibility, time, and innovation). In the second measurement model, a second-order construct 'BDA maturity' is conceptualised (reflecting the first-order factors data generation capability, data integration and management capability, advanced analysis capability, digital analytics capability, data visualisation capability, data-driven culture, big data skills) along with the mediating and dependent variables. Hair et al. (2010) argued that when theoretical support exists to conceptualise a second-order construct, SEM can be used to estimate such a higher-order model provided the model fit is adequate. While the second-order CFA (SOCFA) model is seldom used in academic research, it has many advantages over the first-order factor model (FOCFA) (Bagozzi et al. 1991). First, it is argued that in a second-order CFA model "random error variance and specific variance are not confounded" (p.439). Second, "it presumes that variation in measures will be a linear combination of traits, methods, and error" (Bagozzi et al. 1991, p.439), which is also a shortcoming of first-order CFA (FOCFA) model.

The goodness-of-fit of both the models is assessed using multiple fit indices, given in Table 4.10, as suggested by Hair et al. (2010). The model fit of measurement models is also evaluated using Internal consistency (composite reliability), indicator reliability, convergent validity (average variance extracted), and discriminant validity (Hair et al. 2014).

Table 4.10 Model-Fit Criteria and Acceptable Fit Interpretation

Model-fit Criteria	Interpretation	Acceptable level
CMIN/DF **χ2/degrees of freedom**	Compares obtained X^2 value with tabled value for given *df.*	*<3 is considered as good, but <5 is also acceptable*
Goodness-of-fit index (GFI)	The covariance between the observed variables is measured, similar to multiple R	A value close to 1 is good, but close to 0 indicate no fit.

	square estimation in multiple regression.	
PCLOSE	It measures whether the significance of RMSEA value is <0.5	PCLOSE greater than 0.5 is considered as good. But, less than 0.05 means no model fit.
Adjusted GFI (AGFI)	It is the value adjusted for degrees of freedom *df*	A value close to 0 means no fit; a value close to 1 measure good fit.
Root-mean-square residual (RMR)	The value specifies the closeness of Σ to *S* matrices	Researcher defines level
Standardized RMR (SRMR)	A value of less than .05 indicates a good model fit	< .05
Root-mean-square error of approximation (RMSEA)	Value of .05 to .08 indicate close fit	.05 to .08
Tucker–Lewis Index (TLI)	Value close to .90 or .95 reflects a good model fit	A value close to 0 means no fit; a value close to 1 measure good fit.
Normed fit index (NFI)	Value close to .90 or .95 reflects a good model fit	A value close to 0 means no fit; a value close to 1 measure good fit.
Parsimony fit index (PNFI)	Compares values in alternative models	A value close to 0 means no fit; a value close to 1 measure good fit.
Comparative fit index (CFI)		≥ .95 for acceptance

Source: (Schumacker and Lomax 2010; Schreiber et al. 2006)

Once the model fit is estimated, the models are tested for common method bias (which measures the measurement errors caused by methodological bias Podsakoff et al. 2003). Further, measurement invariance is assessed in this study for any noticeable between-group differences. Both the differences in terms of the strength of association between observed and latent variables (metric variance) and the differences in terms of varying patterns of factor loadings between-groups (Dimitrov 2010). As data is collected from a population which has multi-groups such as small, medium and large in this study, the between-groups differences have to be investigated. So a multi-group moderation test is conducted and the 'critical ratios' are assessed to find the presence of differences in the multiple groups in the data (Milfont and Fischer 2010).

4.12.2 Structural Equation Modelling

In the previous section, the development of measurement models and model fit assessment are discussed. Consequently, a structural model is developed in this research to test the hypotheses with the quantitative data collected using the survey method. SEM is used in this research to address the following research questions and test the list of hypotheses proposed in chapter 3.

1. What is the relationship between Big Data Analytics capability maturity and firm performance (operational and innovation)?
2. What is the role of absorptive capacity in the relationship between BDA capability maturity and firm performance (operational and innovation)?
3. What is the role of data and information quality in the relationship between BDA capability maturity and firm performance (operational and innovation)?
4. What is the role of supply chain analytics capability on the relationship between BDA capability maturity and firm performance (operational and innovation)?

Accordingly, the theoretical model investigated in this study involves several concepts such as BDA, absorptive capacity, and performances, with first-order and second-order latent factors. The research model in this study has three mediating variables of interest, i.e. ACAP, DIQ, and SCA capability. SEM techniques are argued to be more robust than multiple regression to test mediation between latent variables (Pearl 2012) and recognised as a suitable method for this study. A detailed overview of structural equation modelling (SEM) and the rationale for using it in this research is presented in this section.

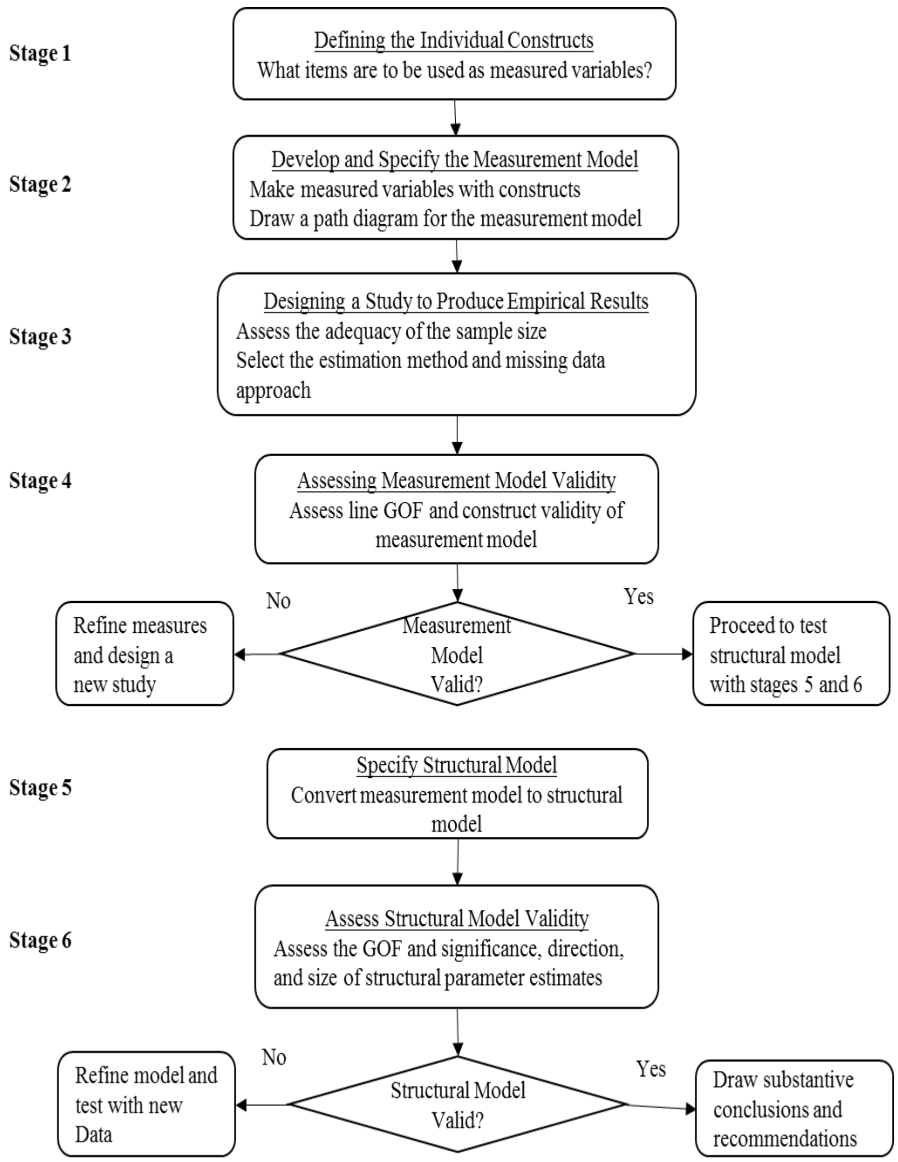


Figure 4.12 Six-Stage Process for Structural Equation Modelling

Source: (Hair et al. 2010)

According to Byrne (2010, p. 3), Structural Equation Modelling (SEM) is “a statistical methodology that takes a confirmatory (i.e. hypothesis-testing) approach to the analysis of a structural theory bearing on some phenomenon.” SEM is an assortment of statistical techniques that can be used to examine a set of relationships between one or more independent variables (IV) and one or more dependent variables (DV) (Tabachnick and Fidell 2012). Both the IV and DV can either be discrete or continuous variables. Structural equation modelling is related to more familiar statistical methods such as Analysis of Variance (ANOVA) and multiple regression (MR) (Hoyle 2012). It is often referred to by various names such as causal modelling, latent variable modelling, path modelling, and analysis of covariance structures

(Tabachnick and Fidell 2012). There are a few differences between SEM and other multivariate techniques (Bentler 2010; Bagozzi and Yi 2012). First, it allows the estimation of the causal relationship between a series of separate but interdependent variables. Second, the observed variable of a latent construct contains measurement errors, but the latent construct is free from such measurement errors either random or systematic errors (Bollen 1989). Bagozzi and Yi (2012) have stated its philosophical foundation: “SEMs provide a useful forum for sense-making and in so doing link philosophy of science criteria to theoretical and empirical research.” SEM is an advanced technique compared to other statistical techniques such as ANOVA and multiple regression (Bollen and Pearl 2013) because SEM can analyse both observed as well as latent variables, but techniques like regression and ANOVA can only analyse observed variables.

SEM is considered as an extension of factor and regression analysis. There are two types of SEM widely used in literature: 1) Covariance Based (CB) – SEM, 2) Partial Least Square (PLS)-SEM. CB-SEM is used to test theories (either confirm or reject) empirically. Whereas, PLS-SEM is mainly used to develop theories and can be used in exploratory research (Hair et al. 2014).

SEM analysis involves an evaluation of both the measurement model and the structural path model. While the measurement model “relates the variables to the constructs”, the structural path model “relates the constructs to other constructs” (Iacobucci 2009). Estimation of the measurement model is literally called as confirmatory factor analysis, which intends to validate or confirm the latent construct and their fit with measurement items (discussed in section 4.12.1.2 Confirmatory factor analysis). CFA is an integral part of SEM as it allows the inclusion of the latent construct and to generate composite variables (Hair et al. 2010). The visual representation of a structural model (a simple mediation model) is given in Figure 4.13.

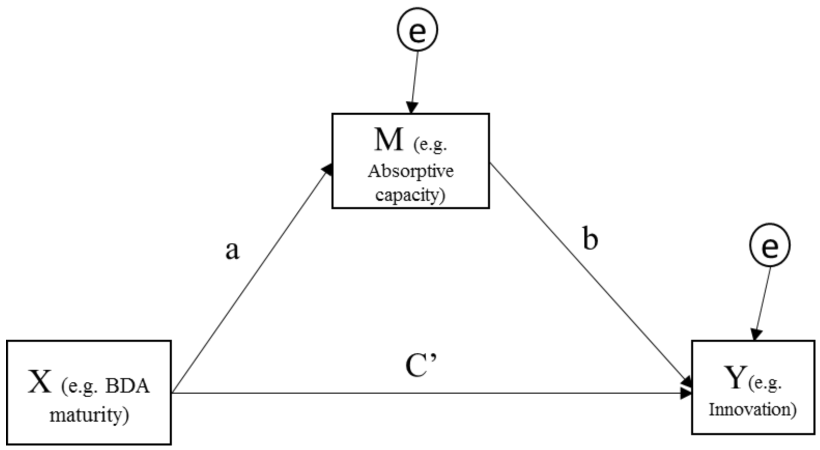


Figure 4.13 Simple three-variable structural path analysis model

Moreover, the six-stage process commonly used for structural equation modelling is adopted in this study (Figure 4.12). Accordingly, EFA is used to identify items that measure the latent variables of the conceptual model (stage 1). Then, the measurement model is developed and validated (stages 2-4). The measurement model is converted into a structural model / path model (stage 5). The path model represents the set of hypothesised relationships between the exogenous and endogenous variables. In this research, the variables used in the path model are called 'latent' as they are observed directly. For instance, in this research, ACAP is a latent variable observed by 12 observed variables. Moreover, as shown in Figure 4.13, a residual or error term is introduced to measure the causal effect between these social variables devoid of external interference. In the same way as the validation of the measurement models, the 'goodness-of-fit' (GOF) of the structural model is evaluated to ascertain model fit (stage 6), the path model is adjusted if necessary. Once the structural model is deemed to valid, the substantive conclusions are drawn from the analysis concerning the hypothesis (Bollen and Pearl 2013).

Apart from testing for direct relationships, SEM can be used to test the mediating effect of latent variables such as ACAP, DIQ and SCA capability on the relationship between BDA maturity and operational and innovation performance. According to Preacher and Hayes (2008, p. 879), "mediation hypotheses posit how, or by what means, an independent variable (X) affects a dependent variable (Y) through one or more potential intervening variables, or mediators (M)." There are three methods available to calculate the size of the mediation effect; 1) causal steps approach, 2) product-of-coefficient approach, 3) difference-in-coefficient (Preacher and Hayes 2008a). As suggested by Hayes (2013), the indirect effect is tested using product-of-coefficient tests with bootstrapping method. The bootstrapping process involves taking a sample ‘n’ number of cases from the original sample with replacement, not just once but multiple times. The graphical representation of the bootstrapping process is shown in Figure 4.14. As shown in Figure 4.14, as a result of extracting ‘n’ number of bootstrap samples, the sampling distribution tends to be normal, and hence it is considered as a more preferred method (Preacher and Hayes 2008a). Moreover, the results of simulation studies reveal that bootstrapping is a robust method for obtaining confidence limits of indirect effects (Preacher and Hayes 2008a; Briggs et al. 2012). Consequently, in this research, the mediation effects are tested in SPSS AMOS software using 5,000 bootstrap samples with a 95% confidence interval.

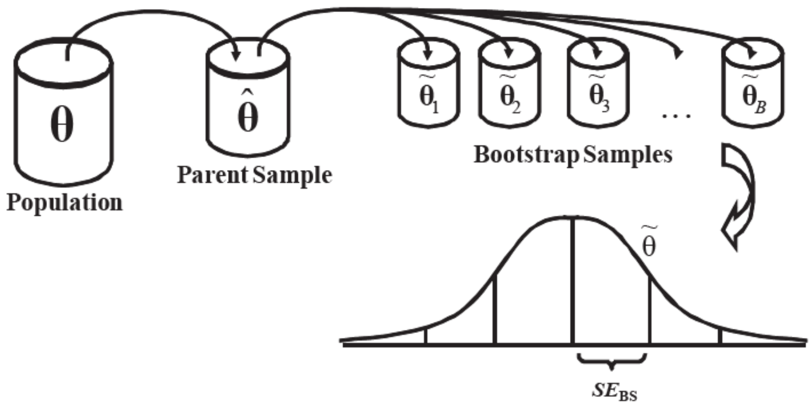


Figure 4.14 Illustration of Bootstrapping process

Source: (Hoyle 2012)

The sample size is an important criterion in SEM analysis as it can affect the stability of the parameter estimates (Schreiber et al. 2006). However, with regards to the sample size requirement for SEM analysis, there is a lack of agreement among researchers. Bentler (1990) recommended a 5:1 ratio of sample size to the number of free parameters estimated. (Schreiber et al. 2006) argued that the 10:1 ratio is more suitable, though the sample size requirement depends on data normality as well. In case the normality of the data is affected, more data samples are required, if not the requirement can be flexible. Iacobucci (2009) argued that a large sample size is required only in situations where the constructs in the structural model do not discriminate against each other well or the estimated effects of predictors on criterion are low. However, as a rule of thumb, researchers have suggested a sample of 200 and above to be considered as acceptable for SEM (Loehlin and Beaujean 2017; Sivo et al. 2006).

In summary, the ability of SEM to analyse a model with both latent and observed variables makes it an appropriate technique for this research, compared to traditional approached such as multiple regression and Analysis of Variance (ANOVA) (Kline 2015; Hair et al. 2014). Besides, many recent studies in operations research and information systems discipline have widely adopted SEM techniques to analyse survey data and test hypothesis (Chae, Yang, Olson, et al. 2014; Bruque Camara et al. 2015).

4.12.3 Cluster analysis

In the previous section, the rationale behind the use of SEM to verify the conceptual model of this study is discussed. This section presents the rationale behind the use of the cluster analysis technique to answer the following research question.

1. Does BDA adoption extend the digital divide between SMEs and Large organisations in the UK?

Clustering is an unsupervised learning method, which is one of the widely used techniques of data mining. "Cluster analysis is the process of finding "natural" groupings by grouping "similar" (based on some similarity measure) objects together" (Dy and Brodley 2004, p.1). Cluster analysis is mainly done to identify natural grouping(s) of points, objects, or patterns (Jain 2010). Clustering algorithms do the partitioning of data into a definite number of clusters, which can be denoted as groups, categories or subsets (Xu and Wunsch 2005). Moreover, clustering algorithms are widely used in academic research to address the digital divide. For instance, Ayanso et al. (2010) used cluster analysis to explore the digital divide between countries across the globe based on their information and communication technologies capabilities. Similarly, Cruz-Jesus et al. (2012) used hierarchical and K-means cluster analysis to assess the digital divide across the European Union.

However, clustering is an exploratory technique and its outcome is subject to several factors such the number of clusters chosen a priori and clustering parameters. So, in order to identify an optimal clustering solution, as objectively as possible, it is beneficial if experiments can be conducted with different clustering algorithms, with varying numbers of clusters, and to validate each clustering solution to find the best possible number of clusters. There are large number of clustering algorithms developed in the literature (Kaufman and Rousseeuw 1990). However, considering the scope of this study and the research objective, only the following three clustering algorithms, Hierarchical, K-medoids and Latent Class Analysis (LCA) are used in this research. These clustering algorithms are generally considered as robust (Arunachalam and Kumar 2018), and especially LCA is argued to be suitable for ordinal data sets. Taking this approach allows the researcher to combine the advantages of these clustering algorithms to find an optimal solution. A brief overview of the foundation and implementation procedures of these clustering algorithms is provided in the next section. The implementation of these clustering algorithms is performed using R studio, as it allows to custom code and modify parameters of the algorithms suitable for the data. Some, clustering algorithms require the usage of a dissimilarity matrix as an input function, and the dissimilarity matrices used in this study are given in Chapter 5.

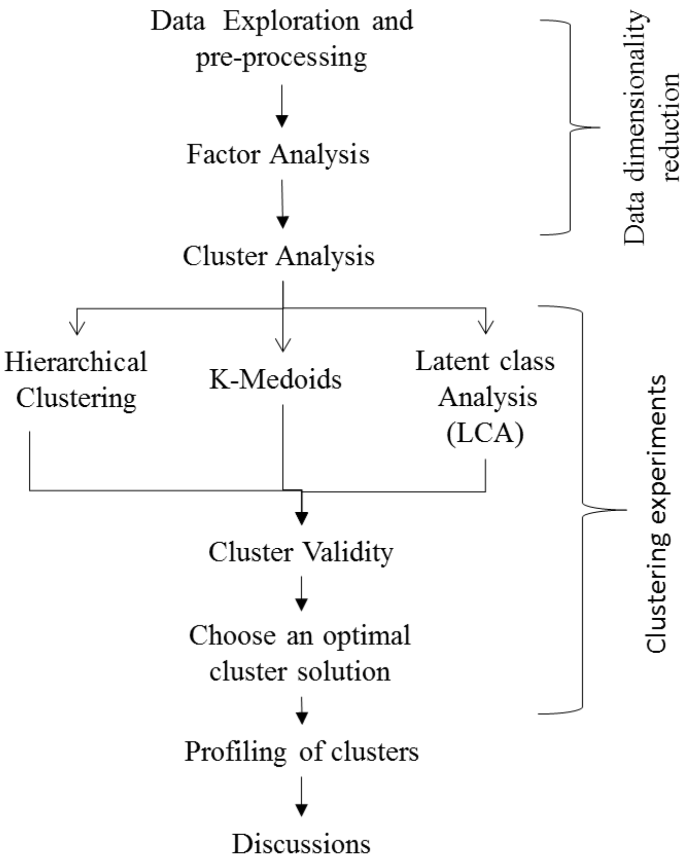


Figure 4.15 Design of cluster analysis experiments

4.12.3.1 Hierarchical Cluster Analysis

Hierarchical cluster analysis (HCA) is one of the most commonly applied technique in management research (Eusébio et al. 2017; Casabayó et al. 2014; Dolničar 2003; Brusco et al. 2017). According to Hair et al. (2010, p. 439), HCA involves "a series of n – 1 clustering decisions (where n equals the number of observations) that combine observations into a hierarchy or a treelike structure". There are two basic types of HCA procedures: 1. Agglomerative (in which the procedure starts with treating each observation as a cluster, and successively the clusters that are similar are joined together to form large clusters), 2. Divisive (in which the procedure starts with treating all the observation as belonging to one single cluster. and are successively divided into smaller clusters). The step-by-step procedure of agglomerative HCA is given in Figure 4.16.

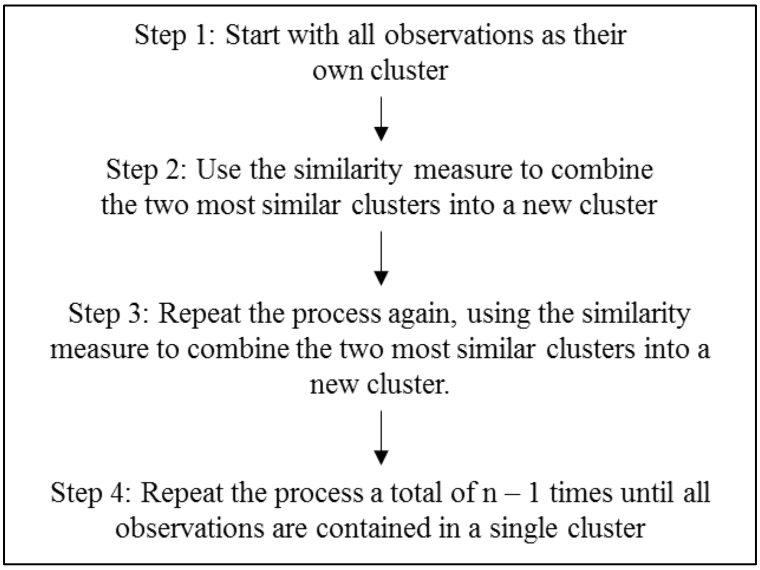


Figure 4.16 Procedure of agglomerative HCA

Sources: adapted from (Hair et al. 2010)

Moreover, some of the popular agglomerative HCAs are: (a) Ward's method, (b) centroid method, (c) average linkage, (d) complete-linkage, and (e) single-linkage (Bridges Jr 1966). A detailed discussion of these variants of HCA can be found in Hair et al. (2010), but the fundamental procedure is analogous to the steps illustrated in Figure 4.16. Consequently, the implementation of these five variants of HCA is carried out in 'R studio' software using a package called 'cluster' (Kaufman and Rousseeuw 1990).

4.12.3.2 K-medoids Clustering

The K-Means algorithm is the most widely used algorithm in practice, even though it is introduced 50 years ago (Jain, 2010). According to Mingoti and Lima (2006), the algorithm starts with assigning K initial seeds of clustering, exactly one seed for each cluster. Euclidean distance is measured, and all of the *n* objects will be compared with the seeds based on the Euclidean distance and the objects which are close to a cluster seed will be assigned to it. Although there are a lot of pros and cons of the K-means algorithm, its applicability for the ordinal data types is highly debatable. Yet, it is widely used to analyse survey data which is either ordinal in nature or mixed data types. However, a more robust clustering method based on K-means is introduced which is called K-medoids algorithm. Unlike K-means, K-medoids uses median values as a cluster centre. Also, according to Kaufman and Rousseeuw (1990), "k-Medoids minimizes a sum of dissimilarities instead of a sum of squared Euclidean distances". K-medoids is a non-hierarchical procedure, commonly referred to as partition-based clustering. The objective function of K-medoids is given below, and Figure 4.17 illustrates the steps involved in K-medoids clustering. In this research, several packages such as 'cluster', 'clustoplot' 'and daisy' are used to implement the k-medoids clustering and visualise the clustering solutions.

Objective function = S $d(i, mv_I)$

Where, 'i' – denotes objects in 'v_I 'cluster, mv_I - is the medoid.

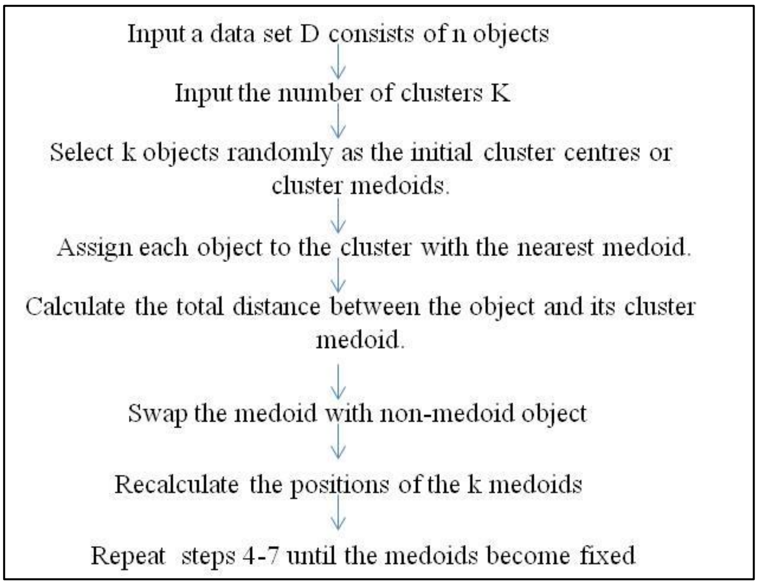


Figure 4.17 Steps involved in k-Medoids clustering

Source : (Cao and Yang, 2010)

4.12.3.3 Latent Class analysis

Latent class analysis (LCA) involves the building of latent classes which are unobserved latent subclasses or clusters of cases (Collins and Lanza 2010). The LCA is different from the standard cluster algorithms such as HCA and K-medoids. LCA is a model-based clustering approach and it assumes that “the data are generated by a mixture of underlying probability distributions” (Hagenaars and McCutcheon 2002, p.90). The objective of LCA is to maximise the log-likelihood function to solve clustering problems. Unlike, standard cluster analysis methods, LCA does probability-based classification and it is possible to include variables of ordinal scales within the model (Hennig et al. 2015). There are several names used to describe LCA such as the Maximum-likelihood approach to clustering, probabilistic clustering, and latent class cluster analysis (Hagenaars and McCutcheon 2002; Bartholomew et al. 2011). So, for LCA clustering, all the individual variables measured (in Likert scale) are used as inputs, not the factor score.

Hagenaars and McCutcheon (2002) have discussed the two possible methods to estimate model parameter: 1. Maximum-likelihood (ML) method, 2. Maximum-Posterior (MAP) method. In this research, the ‘poLCA’ package in R studio is used to implement LCA. The poLCA package takes advantage of the expectation-maximization (EM) algorithm to find the ML estimates of the latent class models (Linzer and Lewis 2011). Accordingly, the algorithm aims to maximise the log-likelihood function given by Equation 1. The iterative steps involved in maximising the objective function are given in Figure 4.18. In LCA clustering, the results may vary due to random initialisation of this algorithm. So, several iterations are necessary to overcome the possibility of identifying local minimum instead of the global minimum (Haughton et al. 2009).

$$ln\ L = \sum_{i=1}^{N} ln \ \sum_{r=1}^{R} P_r \prod_{j=1}^{J} \prod_{k=1}^{K_j} (\pi_{jrk})^{Y_{ijk}} \qquad \text{Equation 1}$$

Step 1: Begin with arbitrary initial values of P_r and π_{jrk}, and label them P_r^{old} and π_{jrk}^{old}.

↓

Step 2: Expectation step, calculate the "missing" class membership probabilities

↓

Step 3: Maximization step, up-date the parameter estimates by maximizing the log-likelihood function.

↓

Repeats these steps, assigning the new to the old, until the overall log-likelihood reaches a maximum and ceases to increment beyond some arbitrarily small value.

Figure 4.18 Steps involved in LCA

Source: adapted from (Linzer and Lewis 2011)

4.12.3.5 Cluster validity measures

It is argued that even after a cautious analysis of a data set to obtain the final cluster solution, there is no guarantee of having attained a meaningful cluster (Punj and Stewart 1995). Clustering algorithms find a cluster solution even if there are no natural groupings hidden in the data set. However, tests can be applied to validate and determine the appropriateness of clustering solutions (Punj and Stewart 1995).

Clustering algorithms are sensitive to initial assumptions and since it is an unsupervised technique, a proper evaluation method is required to assess the clustering results. According to (Pakhira et al. 2004), two important aspects that need to be addressed while performing cluster analysis are: 1) to determine the actual number of clusters present in the data, 2) to validate the goodness of clustering. Cluster validity measures provide an effective way to measure the quality of clustering algorithms to find natural groups of the data set (Gan et al. 2007). There are several cluster validity indexes described in the academic literature (Maulik and Bandyopadhyay 2002), (Wu and Yang 2005), (Pakhira et al. 2004). Cluster validity indexes can be broadly classified into external and internal validity indexes. External cluster validation uses information not inherent to the data, whereas internal validation relies on only the information within the data (Liu et al. 2010). Liu et al. (2010) have described that external validation requires class label information to find the best clustering algorithm, but internal validation can find both the best clustering algorithm as well as an optimal number of clusters in the data with the information inherent to the data.

Some of the significant internal cluster validity indices widely used in the academic literature are the Dunn index (Dunn 1973), the Davies-Bouldin Index (DB) (Davies and Bouldin 1979), the Xie-Beni (XB) index (Xie and Beni 1991), the Silhouette index (Rousseeuw 1987), etc. These internal cluster validity indexes are applied to measure two aspects of clustering, i.e. compactness (a measure of closeness of objects within the cluster), and separation (a measure of well-separation of clusters). The silhouette index is based on measuring the pairwise difference of within- and between-cluster distances and a large value indicates good clustering. Xie and Beni (1991) have described that the XB index can be used to measure inter-cluster separation (“calculates the minimum square distance between cluster centers”) and intra-cluster compactness (“calculates the mean square distance between each object and its cluster center”). Similarly, Dunn’s Index can be used to calculate the “inter-cluster separation as the minimum pairwise distance between objects in different clusters and the intra-cluster compactness as the maximum diameter among all clusters” (Dunn 1973).

The DB index can be calculated using the formula below and illustrated in Figure 4.19.

$$DBI = \frac{1}{k}\sum_{i=0}^{k} \max \left(\sum_{j=0.i\neq j}^{k} \frac{avg(i)+avg(j)}{distance(c(i),c(j))}\right) \tag{6}$$

Source : (Momin, 2006)

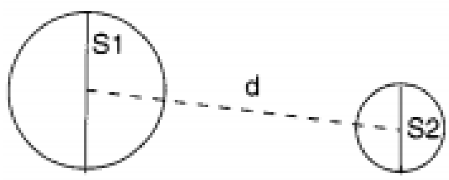


Figure 4.19 Illustration of DB index

Pakhira et al. (2004) have reinforced that the Dunn Index and the DB Index are well suited for validating crisp clusters, and the XB index for fuzzy clusters. Hence the choice of cluster validity index is also dependent on the type of clustering algorithm used in the experiment. Moreover, for validating LCA model-based clustering solutions, two parsimony criteria are used in this research: 1. Bayesian Information Criterion (BIC) (Schwarz 1978), 2. Akaike Information Criterion (AIC) (Akaike 1973).

In this study, the validity of the clusters is measured using above mentioned validity indexes. In addition, once the valid clusters are identified, an optimal clustering solution is chosen, and profiling of clusters is performed to make the association between target classes and organisational characteristics such as firm size and annual turnover.

4.13 Summary of methodology chapter

Research methodology is an important part of this research. This chapter has provided detailed discussions on the rationale behind the strategic selection of a mono-method quantitative study. The research design of this study is informed by both philosophical and practical considerations, along with the research aims. Table 4.11 represents the summary of the research methodology adopted in this study. The research philosophy of this study is critical realism, and the theory development approach is deductive in nature. Further, a web-based self-administered questionnaire survey is used to collect quantitative data. Two types of data analysis techniques are used in this research: 1) Structural Equation Modelling (SEM), 2) Cluster analysis.

Table 4.11 Summary of research methodology of this study

Key criteria	Explanation
Research philosophy	Critical Realism
Theory development approach	Deductive approach
Research method	Mono-Method (Questionnaire survey)
Data collection method	Quantitative data: Web-based self-administered questionnaire survey
Potential participants	Senior executive such as Chief Executive officer (CEO), Chief information officer (CIO), IT managers/ Big Data professionals, supply chain professionals Woking in the manufacturing companies of the UK.
Sampling strategy	Random sampling
Data analysis method	**Quantitative data analysis:** a) Exploratory factor analysis (EFA)- using SPSS b) Confirmatory factor analysis (CFA)- using SPSS AMOS c) Structural Equation Modelling (SEM)- using SPSS AMOS d) Cluster Analysis- R Studio (open source software)

Chapter 5 Data Analysis and Findings

5.1 Chapter Introduction

This chapter presents: a) the preliminary data screening and data preparation process, b) findings from Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA), c) findings from the path analysis using Structural Equation Modelling (SEM), which is used to test the hypothesis. In addition, findings of cluster analysis is also presented, which is used to explore the phenomenon of the digital divide caused by the disparity in the adoption trend of BDA in the UK manufacturing industry.

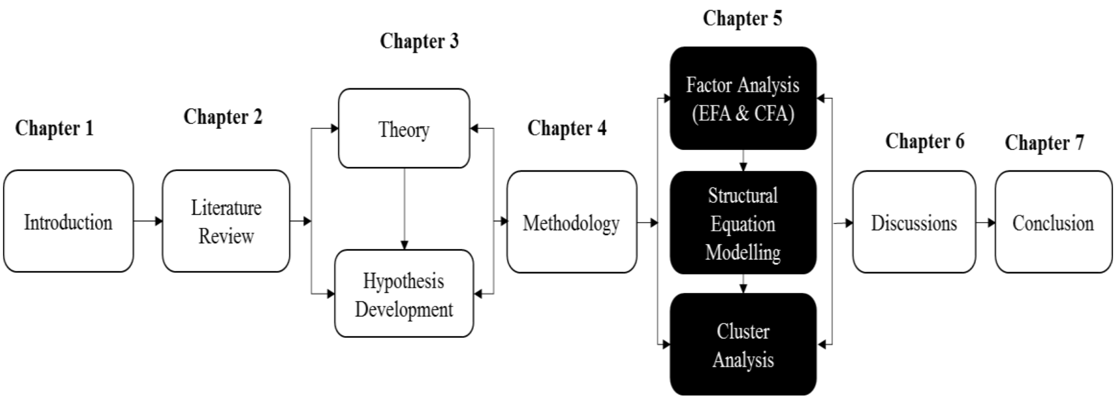


Figure 5.1 Position of the chapter in this thesis

5.2 Preliminary Data Screening

The characteristics, quality and nature of data collected has to be investigated before performing hypotheses testing (Tabachnick and Fidell 2012; Hair et al. 2010). The preliminary examination and assessing the suitability of data to perform inferential analysis, involves three separate tasks, i.e. missing data evaluation, outlier identification and ensuring that the data satisfies all the underlying assumption for multivariate data analysis such as normality (Hair et al. 2010). The consequence of not conforming to these tests and underlying assumptions would lead to the occurrence of Type I or Type II errors. These errors would create problems such as overestimation or underestimation of the effect sizes. This section covers the preliminary data analysis conducted by examining various characteristics such as central tendencies, mean, standard deviation and frequency distribution.

5.2.1 Data Gathering and Data screening

In this research data is collected using respondents identified from Fame Database and other industry sources. The Fame database contains details of 158415 manufacturing companies. After employing the inclusion criteria: 1. Senior executives (CEO, IT and operations), 2. Only one contact from a company, and 3. Only contacts with an email address and a telephone number, a total of 23,608 contacts out of 15,8415 manufacturing companies were left for the sampling. But, the FAME database is constrainted due the lisitng of contacts with unknown number of employees. So, all the contacts without the data on 'number of employees' are excluded from the sampling process. Consequently, with further shortlisting based on availability of the number of employees; 15,425 companies belong to SMEs (i.e. 1 to 250 employees) and 2,148 large (i.e. above 250 employees) manufacturing organisations are identified from the database. Out of 2,148 large companies, only 1,154 contacts of senior executives are with both email address and telephone number. Similarly, out of 15,425 SMEs, there are only 4,898 senior executive contacts with both email addresses and telephone numbers. This process leads to a sample of 6052 valid contacts. In this research, data gathering started in the month of April 2017 and continued until October 2017. In the first phase of data collection, the questionnaire survey link was distributed via Qualtrics to the 2,000 email addresses identified from FAME. However, the Qualtrics system notified that 284 email addresses are bounced back, indicating these email addresses are inactive due to several reasons such as employees retiring or moving to another organisation. Qualtrics is used to set periodic email reminders and follow up telephone calls were made to encourage non-respondents. However, there are several cases (approximately 50), in which the researcher was emailed back by the participants that they cannot participate in the research due to the following reasons: 1) the company's policy prohibits them to participate in external surveys and provide information about the organisations' resources and performance, 2) no knowledge of the research topic, 3) simply not interested in this research. Subsequently, the email addresses of those participants not interested in the survey are manually removed from the database. This process permitted us to not send unnecessary reminders to the participants who had expressed concern over participating in the survey. Moreover, there are respondents who expressed doubts about the questionnaire and wanted to know more about big data analytics, and the value this research would bring to their company and to the industry in general. Further, the questionnaire link was sent to industrial contacts known to the researcher and supervisors. In addition, manufacturing companies listed in the website of organisations such as Advanced Manufacturing Research Centre (AMRC), Doncaster Chamber of Commerce and

Barnsley & Rotherham Chamber of Commerce are contacted via telephone to source email addresses of potential participants. FAME database was also used as a reference tool to identify and contact manufacturing companies via telephone and source email addresses. Then, the survey link was sent to all the email addresses sourced through the telephone process on ad-hoc basis. Besides, the link was also sent to industry professionals who participated in ‘Big Data for SMEs’ workshop organised at the University Of Sheffield Management School. In addition, the online questionnaire link was sent to professional contacts (senior executives working in the UK manufacturing sector) identified via ‘Linkedin’. The practice of collecting survey responses from participants identified via professional network such as ‘Linkedin’ is evident in recent research works such as Gupta and George (2016), Dubey et al. (2015), Belekoukias et al (2014) and DeGroote and Marx (2013). In total 7 scheduled email reminders (4 weeks apart) and a final reminder before the closing of the survey were sent during the entire process of data gathering. In total, the questionnaire link was sent to 2174 respondents via Qualitrics and email. However, only a total of 334 submitted responses (including responses with missing values) were received. After rigorous data preprocessing (section 5.2.2), 221 responses were considered as usable for further analysis, a response rate of 10.2%.. The obtained sample size (>200) is considered sufficient for data analysis and hypothesis testing (Hair et al. 2010; Tomarken and Waller 2005). Since the data collection occurred in mutliple waves, a non-response bias test was conducted to asses the characteristics of early and late respondents and identify if there are any potential bias in the data before merging the two sets. A T-test revealed that there is no significant difference between early and late respondents, thus reducing the risk of non-response bias (Armstrong and Overton 1977).

After the closure of the survey, the data gathered was imported into SPSS software and the initial data screening was conducted. The data entry into SPSS software was performed by the researcher and the variables are coded meticulously to avoid transcription errors. Furthermore, in order to perform structural equation modelling in the AMOS software, the pre-processed data file had to be in the SPSS file format. So, the majority of data pre-processing and exploratory factor analysis are conducted using SPSS, and SEM was performed using the AMOS SPSS software. However, cluster analysis was performed using R open-source analytics software as it had several functions to apply various types of clustering algorithms, which are not available in SPSS.

5.2.2 Missing Data Analysis

Missing data is omnipresent in social science research for decades and researchers use several techniques to treat the missing values and ensure data quality. However, Enders (2010) argued that while there are many techniques available to deal with missing values the assumption about the cause of the missing values is susceptible to bias. Little and Rubin (2002) accentuated the importance of recognising the missing-data pattern and missing-data mechanism. Missing-data pattern refers to “which values are observed in the data matrix and which values are missing”, and the missing data mechanism (or mechanisms), refers to “the relationship between missingness and the values of variables in the data matrix” (Little and Rubin 2002, p.4). Moreover, missing values in a dataset can occur in three possible ways i.e., Missing not at random (MNAR), Missing at random (MAT), Missing completely at random (MCAT) (Schminkey et al. 2016). There are several ways to deal with missing data such as listwise deletion, pairwise deletion and data imputation (Tomarken and Waller 2005). However, the choice of missing data treatment depends on the missing-data mechanism. It is essential to determine whether the occurrence of missing values is at random or not. Using SPSS, Little’s MCAR test is performed on the 334 responses and the results suggested that the values are missing completely at random. Consequently, the treatment of the missing values could be conducted, either by imputing with the Expected Maximisation (EM) approach or removed via a list-wise or a pair-wise strategy. However, the imputation technique raises concern over potential bias. So, in this research, list-wise deletion strategy is adopted, and 81 survey responses (incomplete questionnaires) having more than 50% missing values are removed from further analysis.

From the remaining 253, 29 responses were removed due to missing values on key variables (i.e, demographic variables) or if the respondents failed to satisfy the screening criteria, either not belonging to the manufacturing sector or scoring less than three on the Likert scale of 1 to 5 (Not aware at all - Extremely aware) for the screening questions. The screening questions are intended to test the level of adoption and the respondents’ knowledge and awareness of three aspects related to their organisation such as BDA resources, supply chain environment and decision-making processes. In addition, 3 responses are removed based on the location of the participants' organisation as these three cases mentioned that while their sales office is in the UK, their manufacturing operations are in some other European country. Finally, 221 valid responses remained after rigorous pre-processing and this is considered sufficient to perform structural equation modelling (Hair et al. 2010; Tomarken and Waller 2005).

5.3 Demographics and Descriptive Analysis

In the survey, the respondents are asked to provide various demographic information such as the number of employees, annual turnover, the manufacturing sector, years of experience in the organisation and the location of the organisation. In this section, the descriptive statistics of the demographic variables are analysed to give an overview of the sample characteristics.

5.3.1 Respondents demographic profile

In Table 5.1, it is evident that more than 50 per cent of respondents have at least 5 years or more work experience in the participating organisations. Since senior executives are the primary target of the survey and our sampling strategy is well devised, most of the respondents participated in the survey can be considered as well informed about the organisations' decision-making processes and performances. Especially, investments in new technologies such BDA is an un-programmed strategic decision (Boddy 2014) and the senior executives of an organisation hold significant responsibility for monitoring the adoption and impact of such technologies on performances.

Table 5.1 Respondents experience in the organisation

Categories	Frequency	Per cent
Less than 6 months	3	1.4
6 - 12 months	3	1.4
1 - 5 years	34	15.4
5 - 10 years	64	29.0
10 -15 years	33	14.9
15-20 years	35	15.8
More than 20 years	48	21.7
Total	221	100.0

Similarly, Table 5.2 presents the job title of the respondents. Most of them hold middle level and senior level managerial positions in their respective organisations. In particular, the top contributors to the survey are CEOs, General managers, and senior managers in Information Technology or Big Data Analytics professionals. However, there are a few participants from low-level managerial positions such as IS support and Business Analytics. In these cases, the decision is made to retain them considering their knowledge of BDA, firm performance and the years of experience in the organisations. Overall, based on the respondents 'work experience' and 'job title', the data sample can be considered most suitable for the research context and ensures the authenticity of the responses.

Table 5.2 Respondents job title

Categories	Frequency	Percent
Chief Executive officer	40	18.1
General Manager	38	17.2
Senior Manager (Information Technology/Big Data Analytics/supply chain analytics)	38	17.2
Director	33	14.5
Senior Manager (Product Development/ Operations/ Supply chain/ Logistics /Waste management etc.)	23	10.4
Chief Information Officer	19	8.6
Manager (Information Technology/Big Data Analytics/supply chain analytics)	11	5.0
Manager (Product Development/ Operations/ Supply chain/ Logistics / Waste management etc.)	7	3.5
Business Analyst	1	0.5
Business Manager - Automotive	1	0.5
Chairman	1	0.5
Engineering Manager	1	0.5
Finance Director	1	0.5
Head of Business Development	1	0.5
IS Support	1	0.5
Marketing Director	1	0.5
Owner	1	0.5
President	1	0.5
Technical Manager	1	0.5
Vice President	1	0.5
Total	221	100.0

5.3.2 Organisations demographic profile

The organisations' demographic profile is measured via four different variables such as 'number of employees', annual turnover, manufacturing sector and location of the organisation. As given in Table 5.3, among the 221 organisations the business operation of a high number of organisations participated in the survey falls into the categories of 'Electrical equipment', 'Metals' and 'Food and dairy products'. However, considering the percentage figures, no one sector dominates the sample composition, as it is spread across various sectors of the manufacturing industry. This indicates the generalisability of the findings to all the sectors in the UK manufacturing industry.

Table 5.3 Manufacturing sector

Categories	Frequency	Per cent
Electrical equipment	34	15.4
Metals	31	14.0
Food and dairy products	22	10.0
Computer, electronic and optical products	21	9.5
Chemicals	16	7.2
Non-metallic mineral products	13	5.9
Rubber and plastic products	13	5.9
Automobiles	11	5.0
Machinery and equipment	9	4.1
Textiles	9	4.1
Pharmaceuticals	8	3.6
Beverages	4	1.8
Printing and reproduction of recorded media	4	1.8
3D printing	3	1.4
Coke and refined petroleum products	3	1.4
Papers and leathers	3	1.4
Physical Security products	2	0.9
Remanufacturing / Recycling & Waste	2	0.9
All that handle or process bulk solids	1	0.5
Construction fixings and Security products	1	0.5
Disability equipment / wheels	1	0.5
Filtration equipment	1	0.5
Fishing products	1	0.5
Furniture	1	0.5
Highway Products	1	0.5
Manufacturing machinery coil winding	1	0.5
Play & Sports Equipment	1	0.5
Preservatives manufacturer	1	0.5
Turned and milled parts	1	0.5
Various	1	0.5
Wood Products	1	0.5
Total	221	100.0

Moreover, according to the office for national statistics (Chapman 2017), the spatial distribution of manufacturing firms in the UK is highly concentrated in the following regions: North East, Yorkshire and The Humber, Wales, West Midlands and East Midlands. As shown in

Table 5.4, the location of the participating organisations roughly concurs with the national statistics. Still, there are a few organisations who indicated that they operate in various locations in the UK and they are categorised into 'Multiple locations in the UK'. Reflecting the governmental statistics (Chapman 2017), there are only a few organisations that participated from London and the southern region.

Table 5.4 Location of participating organisations

Categories	Frequency	Per cent
West Midlands	31	14.0
Multiple locations in the UK	27	12.2
South East England	26	11.8
East Midlands	25	11.3
Yorkshire & Humber	24	10.9
North West England	21	9.5
South West England	16	7.2
North East England	15	6.8
Scotland	11	5.0
Wales	11	5.0
East of England	10	4.5
Northern Ireland	2	0.9
London	1	0.5
Midlands: Stoke-on-Trent	1	0.5
Total	221	100.0

Table 5.5 Number of employees and an annual turnover of the participating organisations

Categories	Frequency	Per cent	Categories	Frequency	Per cent
1-9 employees	10	4.5	<= 2Million	35	15.8
10-49 employees	57	25.8	2 - 10 Million	58	26.2
50-99 employees	27	12.2	10 - 25 million	39	17.6
100-249 employees	59	26.7	25 - 50 million	39	17.6
250-500 employees	29	13.1	> 50 million	50	22.6
More than 500 employees	39	17.6	10 - 25 million	39	17.6
Total	221	100.0	Total	221	100.0

Finally, the descriptive analysis of the number of employees and the annual turnover of the participating organisation are given in Table 5.5, indicating the size of the firms. The size of a firm can be measured both by the number of employees working in the organisation and its annual turnover. According to the UK Department of Business, Energy and Industrial Strategy (2017), firms with less than 50 employees are categorised as small, above 50 and below 250 employees are considered as medium-sized firms, and firms with above 250 employees are considered as large firms. Similarly, with regard to annual turnover, firms with less than £10 million are considered as small, between £10 and £50 million are considered as medium size, and above £50 million are defined as a large firm. The profile of sample organisations suggests that, as per both the criteria, the majority of respondents participated in the study are from medium and large firms representing more than 50 % of the sample. However, a considerable number of small firms (approx. 30%) have also participated in the study. The size of the firms and other demographic variables are used in the clustering technique to profile the characteristics of the cluster while exploring the digital divide in the UK.

5.4 Data Inferential Analysis

5.4.1 Assessing Normality

Normality is one of the basic assumptions to perform multivariate analysis, which indicate the symmetric bell-shaped curve distribution of the variables. The default estimation method used in SEM technique in AMOS assumes multivariate normality for continuous variables (Kline 2015). The normality of the data set can be assessed statistically and graphically (Tabachnick and Fidell 2012). The statistical examination involves verifying skewness and kurtosis values whereas the commonly used visualisation techniques are histogram and normal Q-Q plots.

Skewness represents the symmetry and Kurtosis represents that peakedness or flatness of a distribution (Tabachnick and Fidell 2012). If the bell curve is shifted towards the left its positive skew and if shifted towards right its negative skew (Hair et al. 1998). Also, depending upon how peak or flat the bell curve is in comparison to normal distribution kurtosis are classified into 'leptokurtic (peakedness)' and 'platykurtic (flatness)' kurtosis. Theoretically, for a normal distribution, the skewness and kurtosis value must be zero. However, in practice, literature has discussed some acceptable limit for both skewness and kurtosis value. Hair et al. (1998) argued that the distribution of a variable can be considered substantially skewed and non-normal only if the skewness value falls outside the range of -1 to 1. Another rule of thumb is that a skewness value between -0.5 and 0.5 indicates that the distribution is approximately symmetric. Whereas, in the case of kurtosis, the absolute kurtosis value of a variable must be less than the product of the absolute kurtosis value and its standard error to consider it as a normal distribution. The skewness and kurtosis value of all the variables measured in this research is given in Appendix D, which indicate that there is no issue of normality in the data set. All the skewness values are within the range of -1 to +1. In the case of kurtosis statistics, all the values are less than the 0.978 threshold, which is obtained by multiplying the standard error term and the kurtosis value. Subsequently, in this research, the normality of the variables is also assessed by the visual inspection method. The histogram and Q-Q plot of all the variables are generated and investigated. As an example, the histogram and Q-Q plot of item ‘DIM1’ belonging to ‘data integration and management capability’ construct is shown in Figure 5.2 & Figure 5.3. The normal Q-Q plot and detrended Q-Q illustrates the normal distribution of the variable as the points are less deviated from the desired straight line in the Normal Q-Q plot, the outcome is similar for other variables as well. Both statistical and graphical inspection confirms the normality assumptions.

Moreover, when it comes to the impact of sample size on normality assumptions, (Hair et al. 1998, p.70) have stated that “*if the sample size is less than 30 or so, significant departures from normality can have a substantial impact on the results. For sample sizes of 200 or more, however, these same effects may be negligible. As the sample sizes become large, the researcher can be less concerned about nonnormal variables*”. Hence, this research confirms that the data set (which is larger than 200 cases) satisfies the assumption of normality which is mandatory to perform parametric statistical techniques such as SEM.

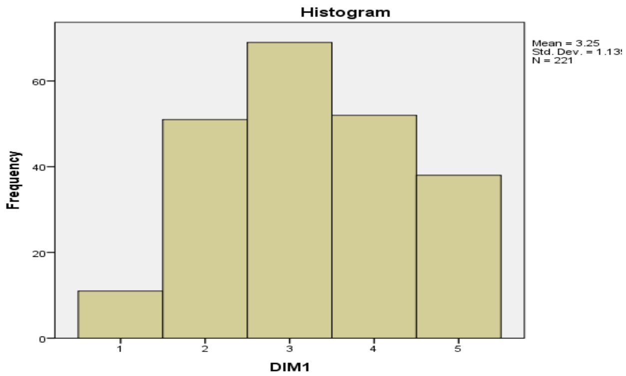


Figure 5.2 Histogram output of DIM1

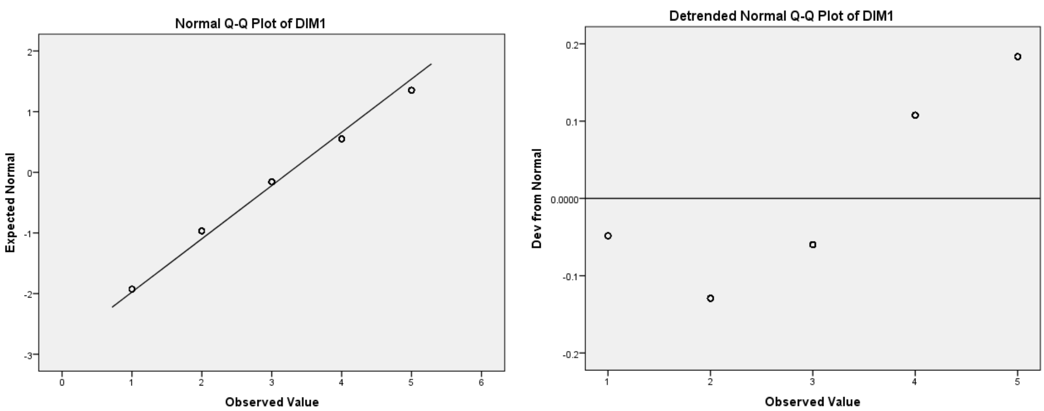


Figure 5.3 Normal and detrended Q-Q plot of DIM1

5.4.2 Outlier identification

Outliers are observations that are extremely different from the rest (Kline 2015). There are two possible outliers: univariate and multivariate outliers. In behavioural research, detecting outliers is a common concern as statistical techniques are sensitive to it (Leys et al. 2018). Tabachnick and Fidell (2012) have stated several reasons for the occurrence of outliers such as incorrect data entry, incorrect specification of imputed missing values, and the presence of cases that are not representative of the target population. Outliers can be identified by investigating the frequency distribution, standard deviation from the mean and Mahalanobis distance, and by using visual inspection techniques such as histogram and box plots. The box plot of the variables is inspected to look for the presence of extreme outliers. There are no significant outliers in the data set. The box plot of 'DIM1' is given in Figure 5.4, suggesting no potential outlier. Moreover, Tabachnick and Fidell (2012) argued that outliers require serious consideration mainly in the case of dichotomous and continuous variables. Since the variables measured in this research are on a five-point Likert scale and as suggested in (Leys et al. 2013), the data set is deemed not to contain any extreme outliers and therefore no mathematical transformation is required.

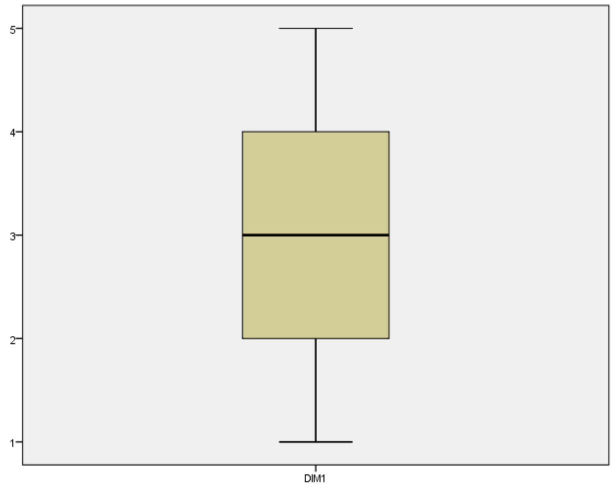


Figure 5.4 Box Plot output of DIM 1

5.4.3 Exploratory Factor Analysis

Factor analysis is a multivariate technique and its main purpose is "to define the underlying structure among the variables in the analysis" (Hair et al. 2010, p.91). The interrelationships between a large set of variables are analysed to find factors, which are a set of variables that are highly correlated. A detailed discussion of the various methods of factor reduction is given in the methodology chapter section

4.11.1. In this section, Exploratory Factor Analysis (EFA) is conducted on variables related to: 1. Organisational capabilities such as BDA capabilities maturity, ACAP, SCA, and DIQ, and 2. Firm performance.

5.4.3.1 Sampling Adequacy Test Results

Before factor analysis, the adequacy of the data sample to perform factor reduction has to be analysed (Kline 2015). In general, the two estimates that are widely used in academic research to assess the ratio of available cases to variables under research are the Kaiser-Meyer-Olkin measure and Bartlett's test of Sphericity. Based on the Kaiser-Meyer-Olkin measure, the cut-off criteria desirable to confirm that the sample data is sufficient to perform a factor analysis is above 0.5. Similarly, Bartlett's Test of Sphericity tests whether the correlation matrix of the variables of choice is significantly different from an identity matrix or not. The required cut-off of Bartlett's Test of Sphericity tests must be significant <0.05 to validate the suitability of the data sample. The results, given in Table 5.6, suggests that both the Kaiser-Meyer-Olkin measure (0.952) and Bartlett's Test of Sphericity (<0.001) have provided satisfactory results to carry out factor reduction analysis.

Table 5.6 KMO and Bartlett's Test results

		EFA on organisational capabilities dimensions	EFA on firm performance dimensions
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		0.957	0.911
Bartlett's Test of Sphericity	Approx. Chi-Square	13,766.910	3,401.014
	df	1,770	171
	Sig.	0.000	0.000

5.4.3.2 Factor Extraction, Communalities and Reliability Test Results

After verifying the sampling adequacy, Exploratory Factor Analysis (EFA) is performed on item scales used to measure the constructs of BDA maturity, Absorptive capacity, data and information quality, supply chain analytics capabilities and firm performance dimensions. As discussed in the methodology section, except 'absorptive capacity', 'data and information quality', 'firm performance dimensions' and few constructs of BDA maturity such as 'big data skills' and 'data-driven culture', items used to measure other constructs are derived from multiple sources. Since this study has adopted item scales from different studies, EFA is carried out to identify scales that are problematic. Moreover, the benefit of EFA is it helps to find hidden factor structure and the inter-relationship between them. While EFA is used to explore the theoretical underpinning of factors, CFA is used to confirm the factor structure (Pallant 2007). Since the factor structure of BDA maturity is not thoroughly investigated in the previous researches, EFA is used to unravel the dimensions of BDA maturity.

To explore and validate the factor structure, the items that load on to non-theorised factors, highly cross-loaded onto other factors and those attained poor loading (<0.5) have to be identified. Moreover, the mean communalities values of all the items are measured and items less than 0.5 are removed from the analysis. The two most prominent factor extraction technique used in academic research is 'Principal Component Analysis (PCA)' and 'Maximum-likelihood' approaches. The details of various EFA techniques are explained in the methodology section 4.12.1. However, the default functionality in AMOS utilises 'Maximum-likelihood' approach to recognise factors of measurement and structural model. Thus, in this study, EFA is performed using 'Maximum-likelihood' extraction approach with 'Promax' rotation using SPSS software. EFA is implemented iteratively with independent, dependent and mediating variables.

First, EFA on variables related to organisational capabilities is conducted. Accordingly, variables that reflect BDA capabilities, DIQ, ACAP and SCA are subjected to factor analysis. As a result, the following items are removed from the further analysis. Items DG1 and DG3 are removed due to cross-loadings, which belongs to data generation capability. Items DIM5 and DIM6 which belong to data integration and management construct is removed due to cross loading onto to several other factors. Item DV4, which belong to the data visualisation construct, is removed due to poor loading (<0.4). Similarly, BS1 and BS2 are removed due to cross loading on other construct.

Table 5.7 Pattern Matrix showing 10 factors of organisation capabilities extracted

Pattern Matrix[a]										
	Factor									
	1	2	3	4	5	6	7	8	9	10
ACAP2	0.915									
ACAP 11	0.880									
ACAP 3	0.876									
ACAP 4	0.873									
ACAP 6	0.868									
ACAP 1	0.864									
ACAP 9	0.856									
ACAP 12	0.854									
ACAP 7	0.850									
ACAP 8	0.843									
ACAP 5	0.789									
ACAP 10	0.788									
SCA6		0.849								
SCA7		0.847								
SCA8		0.843								
SCA10		0.807								
SCA5		0.782								
SCA9		0.766								
SCA2		0.684								
SCA4		0.637								
SCA3		0.635								
SCA1		0.612								
DIM2			0.947							
DIM3			0.896							
DIM1			0.861							
DIM7			0.761							
DIM4RC			0.665							
DIM8			0.663							
DDC5				0.892						
DDC3				0.889						
DDC2				0.809						
DDC4				0.792						
DDC1				0.769						
DIQ4					0.861					
DIQ3					0.848					
DIQ1					0.835					
DIQ5					0.831					
DIQ2					0.749					
AA3						0.959				
AA2						0.917				
AA1						0.743				
AA4						0.560				
AA5						0.555				
BDS4							0.901			

BDS5	0.879			
BDS3	0.840			
BDS6	0.837			
WEBAnalytics		0.915		
SMAnaly		0.877		
TAVanaly		0.682		
DTMining		0.609		
DV2			0.952	
DV3			0.928	
DV1			0.838	
DV5			0.763	
DG5				0.813
DG4				0.759
DG2				0.718
Extraction Method: Maximum Likelihood. Rotation Method: Promax with Kaiser Normalization.				
a. Rotation converged in 7 iterations.				

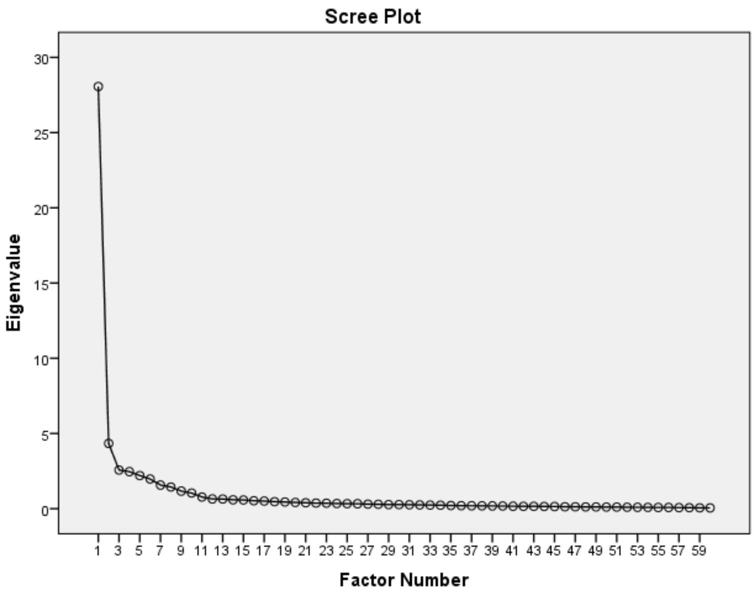


Figure 5.5 Scree plot 10 factors of organisational capabilities

In addition, EFA on firm performance variables are conducted and 5 factors are identified such as time-based performance, product quality, flexibility, cost and innovation. However, variables ‘OP11’ and ‘OP12’ cross-loaded and these variables are removed. The firm performance related factors identified and the rotated pattern matrix indicating the factor loadings are given in Table 5.8. Moreover, by examining the correlation matrix, it is established that the correlation coefficient of all the firm performance variables identified is greater or equal to 0.3.

Overall, there are 15 factors extracted and all the factor loadings are above the threshold of 0.5. Another criterion to determine the factor retention is by inspecting the scree-plot (Costello and Osborne 2005). On the Scree plot, the point where a natural bend occurs and then flattens out indicates the number of factors. Moreover, another criterion suggests only factors with Eigenvalue above 1 can be retained (Tabachnick and Fidell 2012). It can be verified from the 'Eigenvalues', provided in Table 7.6 and Table 7.7 (see Appendix E) which shows that the total variance explained by the factors identified have Eigenvalues above 1. To assess the internal consistency reliability, Cronbach's Alpha (assumes equal indicator loadings) for all the constructs is computed (Hair et al. 2014). The Cronbach's Alpha should be greater than 0.7 cut-off level to consider that the constructs are reliable (Kline 2015). The Cronbach's Alpha values of the constructs are estimated and given in Table 5.9. All the constructs have attained more than 0.8 Cronbach's Alpha value in the reliability test. The scale reliability at the stage of EFA can be considered as a preliminary assessment of reliability. A detailed assessment of construct reliability and validity is performed in section 5.4.4. Hence, based on item loadings, Eigenvalues and preliminary reliability tests, the factors identified from EFA are considered for further analysis and used to build measurement and structural models in the following sections.

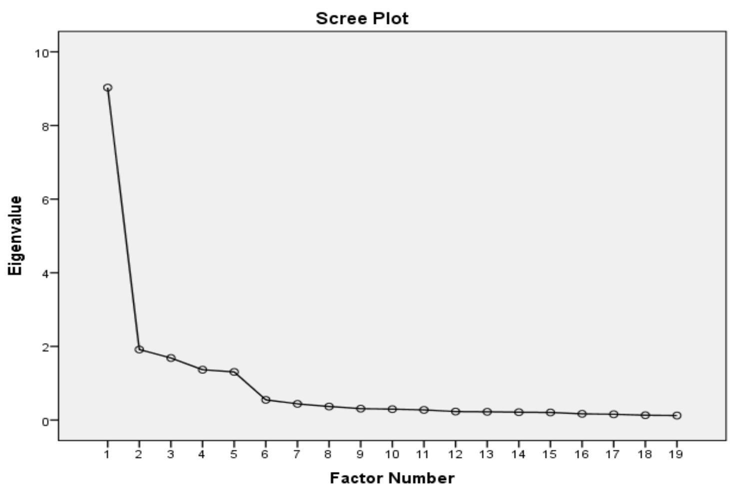


Figure 5.6 Scree plot 5 factors of firm performance

Table 5.8 Pattern Matrix showing 5 factors of firm performance extracted

Pattern Matrix[a]					
	Factor				
	1	2	3	4	5
OP14	0.957				
OP13	0.943				
OP15	0.818				
OP16	0.713				
OP6		0.916			
OP5		0.882			
OP4		0.716			
OP7		0.703			
OP3			0.974		
OP2			0.915		
OP1			0.821		
INNOV1				0.914	
INNOV2				0.874	
INNOV3				0.873	
OP8					0.868
OP10					0.826
OP9					0.782
Extraction Method: Maximum Likelihood. Rotation Method: Promax with Kaiser Normalization.					
a. Rotation converged in 6 iterations.					

Table 5.9 15 factors and scale reliability

Factors	**Cronbach's Alpha**	**Number of items**
Absorptive capacity (ACAP)	0.973	12
Data integration and management (DIM)	0.916	6
Supply Chain Analytics (SCA)	0.943	10
Data-Driven Culture (DDC)	0.918	5
Time-based performance	0.932	4
Data and Information Quality (DIQ)	0.937	5
Advanced Analytics (AA)	0.925	5
Digital Analytics (DA)	0.906	4
Big Data Skills (BDS)	0.960	4
Cost performance	0.893	4
Data Visualisation (DV)	0.953	4
Product- Quality	0.926	3
Flexibility	0.880	3
Innovation	0.908	3

Data Generation (DG)	0.905	3

Table 5.10 Skewness and Kurtosis of factors

	N		Skewness	Std. Error of Skewness	Kurtosis	Std. Error of Kurtosis
	Valid	Missing				
ACAP	221	0	-0.582	0.164	0.337	0.326
DIM	221	0	0.308	0.164	-0.486	0.326
SCA	221	0	-0.029	0.164	-0.325	0.326
DDC	221	0	-0.338	0.164	-0.628	0.326
Time-based performance	221	0	-0.626	0.164	0.146	0.326
DIQ	221	0	-0.401	0.164	-0.139	0.326
AA	221	0	-0.635	0.164	-0.287	0.326
DA	221	0	0.247	0.164	-0.523	0.326
BDS	221	0	0.096	0.164	-0.833	0.326
Cost-based performance	221	0	0.113	0.164	-0.400	0.326
DV	221	0	-0.349	0.164	-0.826	0.326
Product quality	221	0	-0.510	0.164	0.295	0.326
Flexibility	221	0	-0.441	0.164	-0.178	0.326
Innovation	221	0	-0.621	0.164	0.099	0.326
DG	221	0	-0.332	0.164	-0.607	0.326

5.4.4 Confirmatory Factor Analysis

Now that the factor structure is identified using the EFA technique, the next stage is to develop a measurement model and asses the validity and reliability of it. For this purpose, the Analysis of Moment Structures (AMOS) software is used to create measurement models. The assumptions about the discriminant and convergent validity are tested by using Confirmatory Factor Analysis (CFA) (Kline 2015). In this research, two different measurement models are created. As the aim of the research is to find the effect of BDA maturity, as well as its individual dimensions, firm performance dimensions are also considered, hence the requirement for developing two separate models. Theoretically, it is conceptualised that BDA maturity is composed of 7 dimensions such as ‘Data Generation’, Data Integration and Management’,‘Advanced Analytics’, ‘Digital Analytics’, ‘Data Visualisations’, ‘Data-Driven Culture’ and ‘Big Data skills’. In measurement model 2, BDA maturity of an organisation is specified as a second-order construct composed of these 7 dimensions. There are a few recent researchers who have conceptualised BDA capability as a second-order construct such as (Gupta and George 2016) and (Wamba et al. 2017). So, a measurement model with second-order BDA maturity constructs, and a measurement model with the dimensions of BDA capabilities specified as a first-order factor is also developed.

After the measurement models are specified according to the theory and EFA, the goodness-of-fit is assessed using the Model fit plugin available in AMOS (Gaskin and Lim, 2017). The goodness-of-fit for both models are found to be adequate after making some minor changes in the model specification. For instance, based on the modification indices, some of the error terms that have high modification indices, are covaried to improve the model fit. The error terms of DIM7→ DIM8, SCA1 → SCA4, SCA1 → SCA2, SCA6 → SCA7, AC11 → AC12, AC10 → AC9, AC5 → AC6, AC1 → AC2 and AC3 → AC2 are covaried as it showed high modification indices – this procedure is suggested in (Kline 2015). In addition, some items (SCA6, SCA7, SCA10, AA4 and AA5) are removed due to poor loading (<0.7) and issues with modification indices. The model fit statistics obtained after covarying the error terms are found to indicate that the model is suitable for further analysis. The model fit of the first-order Measurement Model is Chi-square (χ^2) =3704.479, degrees of freedom (df) =2587, chi-square goodness-of-fit (CMIN/DF) =1.432, comparative fit index (CFI) =0.932, parsimony comparative fit index (PCFI) =0.868, Normed fit index (NFI) =0.806, root mean squared error of approximation (RMSEA)=0.044, Tucker-Lewis Index (TLI)=0.927 and PCLOSE =0.999. Similarly, the model fit of the Second-order Measurement Model is Chi-square (χ^2) =3805.561, degrees of freedom (df) =2649, chi-square goodness-of-fit (CMIN/DF) =1.437, comparative fit index (CFI) =0.929, parsimony comparative fit index (PCFI) =0.887, Normed fit index (NFI) =0.801, root mean squared error of approximation (RMSEA) =0.045, Tucker-Lewis Index (TLI)=0.926 and PCLOSE =0.998. Both the models attained satisfactory results according to the established cut-off criteria (Hu and Bentler 1999; Byrne 2010).

Apart from assessing the model fit of measurement models, the validity and reliability of the measurement models are also investigated using statistical tests such as Composite Reliability (CR), Average Variance Extracted (AVE) and Mean Shared Variance (MSV). In the previous section, Cronbach's Alpha is used to measure scale reliability indicating the internal consistency of factors. While Composite Reliability also indicates internal consistency, but unlike Cronbach's Alpha, "it does not assume equal indicator loadings" (Hair et al. 2014, p.115). Convergent validity indicates how well the items within the same construct are correlated. On the other hand, discriminant validity indicates how well a construct is different from other constructs in the model (Hair et al. 2014). In this section, three different tests are conducted: 1. the latent construct reliability is measured by estimating ‘Composite Reliability (CR)’, 2. The convergent validity is determined by estimating ‘Average Variance Extracted (AVE)’, 3. The discriminant validity is determined by estimating ‘Mean Shared Variance (MSV)’. The convergent and discriminant validity results of the two measurement models are given in Table 5.11 & Table 5.12. The Composite reliability of all the constructs is above the 0.8 threshold (Byrne 2010; Schumacker and Lomax 2010; Bollen 1989). Moreover, MaxR(H) or Maximum reliability is calculated using AMOS tool (Gaskin and Lim 2017), which is generally considered more robust than CR. Both, CR and MaxR(H) indicate that all the constructs in the model are reliable.

Similarly, convergent validity is also assessed. To consider the models as acceptable the AVE of all the constructs has to be above 0.5. The results indicate that all the constructs in both the models have achieved convergent validity. More than 50 % of the variance in the constructs are explained by the items used to measure it. To test discriminant validity, the Fornell–Larcker criterion is used (Hair et al. 2010), which recommends comparing the square root of AVE and the correlation matrix. Discriminant validity is highly significant while testing for mediation, as the mediators have to be dissimilar from the dependent and independent variables (Zhao et al. 2010). Based on Fornell–Larcker criterion, the square root of the AVE has to be “greater than its highest correlation with any other construct” (Hair et al. 2010). From Table 5.11 and Table 5.12, it is evident that it satisfies Fornell–Larcker criterion and each construct measured in the models are highly dissimilar to other constructs, indicating that all the constructs in the models satisfy discriminant validity.

Moreover, the inter-correlations between the variables are verified. As a general rule of thumb, the inter-correlation between any two variables that is higher than 0.8 indicates multicollinearity. From Table 5.11 & Table 5.12, it is evident that multicollinearity is not a problem in this research as none of the intercorrelations is above the 0.8 threshold. Further, the Variance Inflation Factor ($1/(1 - R^2)$ of the contructs is estimated to assess multicollinearity and the values are below the tolerance level (VIF<10) (Kline, 2015).

Table 5.11 Validity and reliability test results of measurement model 1

	CR	AVE	MSV	MaxR(H)	Digital Analytics	Time	Data Driven culture	Advanced Analytics	Big Data skills	Data generation	Cost	Product Quality	Data Visualisation	Innovation	DIM	Flexibility	SCA	ACAP	DIQ
Digital Analytics	0.905	0.704	0.417	0.909	**0.839**														
Time	0.933	0.777	0.425	0.962	0.346	**0.881**													
Data Driven culture	0.921	0.662	0.417	0.975	0.576	0.430	**0.813**												
Advanced Analytics	0.911	0.721	0.430	0.981	0.450	0.396	0.563	**0.849**											
Big Data skills	0.960	0.858	0.533	0.987	0.585	0.422	0.592	0.610	**0.927**										
Data generation	0.906	0.763	0.507	0.988	0.511	0.396	0.495	0.536	0.587	**0.874**									
Cost	0.894	0.679	0.425	0.990	0.459	0.652	0.454	0.384	0.460	0.485	**0.824**								
Product Quality	0.927	0.810	0.368	0.991	0.358	0.484	0.445	0.449	0.450	0.478	0.404	**0.900**							
Data Visualisation	0.953	0.835	0.533	0.992	0.593	0.391	0.646	0.656	0.730	0.639	0.453	0.430	**0.914**						
Innovation	0.909	0.770	0.346	0.993	0.496	0.460	0.588	0.429	0.373	0.461	0.477	0.401	0.438	**0.877**					
Data Integration and management	0.916	0.612	0.524	0.993	0.646	0.368	0.593	0.582	0.724	0.712	0.503	0.391	0.720	0.424	**0.782**				
Flexibility	0.882	0.714	0.340	0.994	0.346	0.583	0.364	0.466	0.436	0.380	0.490	0.514	0.378	0.445	0.375	**0.845**			
SCA	0.937	0.624	0.510	0.994	0.611	0.541	0.588	0.609	0.653	0.561	0.603	0.607	0.627	0.532	0.646	0.512	**0.790**		
ACAP	0.972	0.744	0.510	0.995	0.547	0.518	0.550	0.560	0.568	0.561	0.568	0.591	0.557	0.588	0.518	0.476	0.614	**0.863**	
DIQ	0.939	0.755	0.366	0.996	0.511	0.516	0.567	0.536	0.579	0.476	0.525	0.501	0.592	0.494	0.534	0.463	0.605	0.556	**0.869**

Table 5.12 Validity and reliability test results of measurement model 2

	CR	AVE	MSV	MaxR(H)	DIQ	Time	Cost	Product Quality	Innovation	Flexibility	BDA Maturity	SCA	ACAP
DIQ	0.939	0.756	0.475	0.945	**0.869**								
Time	0.933	0.777	0.424	0.970	0.516	**0.881**							
Cost	0.894	0.679	0.424	0.976	0.525	0.651	**0.824**						
Product Quality	0.927	0.809	0.370	0.982	0.501	0.484	0.404	**0.900**					
Innovation	0.910	0.770	0.346	0.985	0.494	0.460	0.477	0.401	**0.878**				
Flexibility	0.882	0.714	0.340	0.987	0.463	0.583	0.490	0.515	0.445	**0.845**			
BDA Maturity	0.917	0.615	0.604	0.988	0.689	0.493	0.576	0.537	0.561	0.492	**0.784**		
SCA	0.937	0.623	0.604	0.990	0.605	0.541	0.602	0.608	0.532	0.512	0.677	**0.790**	
ACAP	0.972	0.744	0.510	0.993	0.556	0.518	0.568	0.591	0.588	0.476	0.690	0.614	**0.863**

5.4.5 Common Methods Bias Test Results

The occurrence of measurement errors due to methodological bias is a common problem in behavioural research (Podsakoff et al. 2003). The bias may occur due to the use of a single questionnaire, contextual factors and social desirability and vagueness of items and has to be addressed to eliminate potential errors (Podsakoff and Organ 1986). In this research, a single questionnaire is used to collect both the independent variable and dependent variables. There are several ways in which the bias in the dataset can be identified such as Harman's single-factor test and common latent factor method. Harman's single-factor test, suggests that the variance extracted by the first factor in the Exploratory Factor Analysis (EFA) has to be less than 50% to conclude methodological bias does not play a role (Kwon et al. 2014). In this study, the percentage of variance of the first factor identified from the EFA is less than the threshold indicating that there is no bias in the data set. Moreover, the bias is tested using a common latent factor method by inserting a 'dummy variable', called as Common Latent Factor (CLF), into both measurement models discussed in the previous section. The standardised regression weights of measurement models with and without the common latent factor are compared. It is found out that (Appendix D), only two items SCA5 and ACAP1 belonging to supply chain analytics capabilities and absorptive capacity constructs is affected by bias. However, the bias is only slightly above the threshold of 0.2, indicating that there is no potential bias in the dataset.

.

5.4.6 Configural and Metric Invariance Test Results

The data collected in this research are from multiple groups such as SMEs and large organisations. It is necessary to test for the existence of a varying factor structure between the groups and hence subjected to invariance tests. The 4 groups tested include 'firms with a low number of employees' vs 'firms with a large number of employees', 'Firms with low annual turnover' vs. 'Firms with a high annual turnover. Using the data imputation function in AMOS, the latent factors are imputed and used to perform these tests to reduce computational complications. Metric invariance refers to "equal factor loadings across groups" and configural invariance refers to the existence of "fixed model parameters across all groups" (Dimitrov 2010, p.124). To test metric invariance, the chi-square difference between the constrained and unconstrained models is evaluated and the results (given in Table 5.13 & Table 5.14) indicate that the factor structure is consistent irrespective of different groups in the data set. Furthermore, evaluating the measurement models of different subgroups show that the model fit is adequate for a low number of employees vs. high number of employees' subgroups [fit indices: χ^2 = 204.389; df = 124; χ^2/df =1.6482; CFI =0.965.; PCFI =0.570; PNFI =0.918; PCLOSE=0.283; RMSEA =0.054; TLI=0.940; GFI =0.903], and also for the low turnover vs. high turnover subgroups [fit indices: χ^2 = 206.924; df = 124; χ^2/df =1.669; CFI =0.964;PCFI =0.570; PNFI =0.542; PCLOSE=0.248; RMSEA =0.055; TLI=0.940; GFI =0.898]. It is argued that if the model fit of an unconstrained model having different subgroups is satisfactory, it indicates configural invariance and the groups are equivalent regarding the factor structure (Milfont and Fischer 2010). The findings suggest that the sample data investigated in this research satisfies the conditions of metric and configural invariance.

Table 5.13 Chi-Squared difference significance results for metric invariance test (low number of employees vs high number of employees)

<table>
<tr><th></th><th>Chi-square</th><th>df</th><th>p-val</th><th>Invariant?</th><th rowspan="5">chi-square and df for unconstrained and constrained models</th></tr>
<tr><td>Overall Model</td><td></td><td></td><td></td><td></td></tr>
<tr><td>Unconstrained</td><td>217</td><td>124</td><td></td><td></td></tr>
<tr><td>Fully constrained</td><td>226</td><td>132</td><td></td><td></td></tr>
<tr><td>Number of groups</td><td></td><td>2</td><td></td><td></td></tr>
<tr><td>Difference</td><td>9</td><td>8</td><td>0.342</td><td>YES</td><td rowspan="2">Groups are not different at the model level, however, they may be different at the path level.</td></tr>
<tr><td colspan="2">Chi-square Thresholds</td><td></td><td></td><td></td></tr>
</table>

Table 5.14 Chi-Squared difference significance results for metric invariance test (low turnover vs high turnover firms)

<table>
<tr><th></th><th>Chi-square</th><th>df</th><th>p-val</th><th>Invariant?</th><th rowspan="5">chi-square and df for unconstrained and constrained models.</th></tr>
<tr><td>Overall Model</td><td></td><td></td><td></td><td></td></tr>
<tr><td>Unconstrained</td><td>206.9</td><td>124</td><td></td><td></td></tr>
<tr><td>Fully constrained</td><td>220.1</td><td>132</td><td></td><td></td></tr>
<tr><td>Number of groups</td><td>low turnover vs high turnover firms</td><td>2</td><td></td><td></td></tr>
<tr><td>Difference</td><td>13.2</td><td>8</td><td>0.105</td><td>YES</td><td rowspan="2">Groups are not different at the model level, however, they may be different at the path level.</td></tr>
<tr><td colspan="2">Chi-square Thresholds</td><td></td><td></td><td></td></tr>
</table>

5.5 Structural Equation Modelling

In this section, the procedure adopted to test the mediation effect of absorptive capacity, data and information quality, and supply chain capability on the relationship between BDA capabilities and firm performance dimensions is discussed. All the hypotheses proposed in chapter 3 (Theoretical development) are verified in this section. Moreover, the direct effect of BDA maturity and its individual dimensions on the firm performance dimensions (operational and innovation) is investigated prior to mediation analysis. A brief outline of the direct and mediation analysis techniques is given here before proceeding with the implementation and presentation of the findings.

Mediation analysis is the process in which the goal is to verify how the independent variable (X) exerts its influence on the dependent variable (Y) wherein one or more mediating (M) or intervening variables are causally linked between X and Y (Hayes 2013). An illustration of a mediation model is given in Figure 5.7. Analysing mediation enriches our understanding of a phenomenon not only by explaining the direct relationship between X and Y but also how does X Influences Y. This makes it possible to comprehend the underlying mechanism of a causal relationship. Mediators explain the 'internal psychological significance' on an observable external physical event (Baron and Kenny 1986).

In total, there are three categories of regression tests used to assess mediation. The three categories are; "tests of causal steps, tests of the difference-in-coefficients, and tests of the product-of-coefficient" (MacKinnon et al. 2002). Baron and Kenny (1986), the pioneers of mediation analysis, have argued that a variable can be considered as a significant mediator depending on the extent to which it is accountable for the relationship between the predictor and the criterion. The authors proposed the casual step approach to test the hypothesis and verify the effect of intervening variables. Based on the causal step approach, Baron and Kenny (1986) indicated that to confirm the presence of a mediation effect, three basic conditions must be fulfilled. First, the exogenous variable (X) must have a significant effect on the mediating variable (M). Second, the mediator variable must have a significant effect on the endogenous variable (Y). Third, if the first two conditions are met, the effect of an exogenous variable on endogenous variable must be significantly reduced when controlled for the mediating variable. This causal step approach also requires the total effect of X on Y to be significant (Chen and East 2016). The level of mediation is assessed as follows: In Figure 5.7, If the direct effect c' become non-significant, then it is full mediation. Whereas, if the direct effect in the presence of a mediator is significant but the effect size is reduced, then it is 'partial mediation'. If there is no change in the direct effect even after the introduction of the intervening variable, then it can be considered as no mediation.

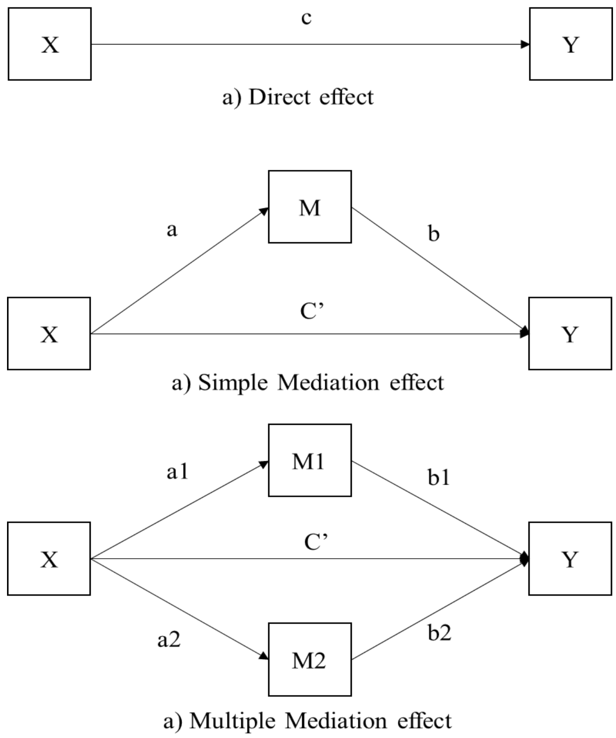


Figure 5.7 An illustration of a) Direct effect, b) Simple mediation effect, c) Multiple mediation effect

In Baron and Kenny's approach, the existence of mediation is based on the outcome of the set of hypotheses. However, this causal-step approach is largely criticised by authors for several reasons (Zhao et al. 2010; Brown 1997; Fritz and Mackinnon 2007). Brown (1997) discussed that testing mediation in sequence by estimating several regression equations ignoring the measurement error in the mediator and assuming that the model is non-recursive i.e. dependent variable cannot cause mediator variable, are the drawbacks of this approach. Zhao et al. (2010) criticised Baron and Kenny (1986)'s classification of mediation effect into full, partial and no mediation as its conception is grounded on a one-dimensional view.

On the other hand, the difference-in-coefficients tests for assessing mediation involves finding the difference between the total effect (c) and the direct effect (C') adjusted for mediating variables (M) and then dividing it by the standard error (Fritz and Mackinnon 2007). Whereas, in product-of-coefficients test, the product of path coefficient X→M (a) and M→Y (b) are divided by the standard error. Hayes (2013) argued that, due to the flaws in the causal step approach and the difference-in-coefficient test, the inferences on indirect effect should be based on the product-of-coefficient tests discussed earlier as it minimizes the number of tests and therefore, the product-of-coefficient approach is adopted in this study.

In the Figure 5.7 a), c represents the total effect of X on Y on the direct effect model. Whereas, in the mediation models, C' represents the direct effect of the predictor on the criterion. The path coefficient between X →M is represented as 'a', and the path coefficient between M → Y is represented as 'b'. Using the product-of-coefficient approach, the indirect effect (a*b) of X on Y via M is calculated by multiplying the path coefficient of X →M and the path coefficient of M→Y. In case of multiple mediation models, the indirect effect of X on Y via M1 is calculated as a1*b1; the indirect effect of X on Y via M2 is calculated as a2*b2, and so on depending on the number of mediators in the model.

In contrast to Baron and Kenny (1986)'s classification of mediation test outcomes into full, partial or no mediation, Zhao et al. (2010) have provided a new typology for mediation and non-mediation effect.

1. Complementary mediation: when the indirect effect (a*b) and direct effect (C') both are significant and if the product coefficient of a,b, C' (a*b*C') is positive.
2. Competitive mediation: when the indirect effect (a*b) and direct effect (C') both are significant and if the product coefficient of a,b, C' (a*b*C') is negative.
3. Indirect-only mediation: when the indirect effect (a*b) is significant but the direct effect (c') on the mediation model is not significant.
4. Direct-only Non-mediation: When the direct (C') in the mediation model is significant, but the indirect effect (a*b) is not significant.
5. No-effect Non-mediation: When both indirect and direct effects are non-significant.

Moreover, the Sobel test and bootstrapping are the two prominent ways used to test the effect of intervening variables based on the product-of-coefficient approach (Bollen and Stinet 1990). Nevertheless, there are three reasons for preferring the bootstrap approach over Sobel tests: 1. Sobel test is based on the assumption of normal distribution of indirect effect which in practice tend to be asymmetric, 2. It is less suitable for small sample sizes (Preacher and Hayes 2008b), 3. Sobel tests use standard error to assess mediation effect, but bootstrapping does not require estimation of the standard error of indirect effect. Besides, the bootstrap method ignores normality assumptions and is generally considered more robust than the Sobel test (Hayes 2013). In this research, the bootstrapping approach is used, and the analysis is carried out by generating 5,000 bootstrap samples and a 95% confidence interval is used to tests the significance of the indirect effect. The outcome of mediation analysis is interpreted as discussed in (Zhao et al. 2010).

5.5.1 Design of the structural model and model fit assessment

A full structural model is developed incorporating the latent factors proven to be valid from the CFA. While developing a complex structural model, as the one presented in this research, the ‘Item Parcel’ approach is generally adopted to simplify the model by reducing the number of indicators of the construct to 3 or 5. A detailed discussion on the ‘Item parcel’ approach is available in (Rocha and Chelladurai 2012), (Meade and Kroustalis 2006) and (Hagtvet and Nasser 2004). However, since there is a mixed opinion about whether to use the item parcel approach among academics and as it is not widely adopted in supply chain researches, the researcher has decided to use the full model without reducing it.

Moreover, our aim is to analyse the impact of individual dimensions of BDA capabilities of an organisation, and as well as the maturity of an organisation on the firm performance. So, two separate structural models are created, one with first-order dimensions of BDA capabilities and another with BDA maturity as a second-order reflective construct. The direct effect of independent variables and mediating variables on the dependent variables is measured before initiating the mediation analysis. However, due to complexity, the effect all the mediators are studied only on the relationship between BDA maturity (second-order construct) and firm performance dimensions. Further, two variables that measure firm size (Number of employees and annual turnover) are included in the structural variable as ‘control variables’. The full structural model is specified as given in Figure 5.8. The goodness-of-fit statistics of structural model is assessed and found satisfactory (Table 5.15). Consequently, as discussed in the previous section, the bootstrapping method is used to find the specific indirect effect of each mediator in the model. The results of direct effect analysis and mediation analysis are presented in the following sections.

Table 5.15 Model fit of the structural model

Measure	Estimate	Threshold	Interpretation
CMIN	4185.055	--	--
DF	2800	--	--
CMIN/DF	1.495	Between 1 and 3	Excellent
CFI	0.918	>0.95	Acceptable
SRMR	0.081	<0.08	Acceptable
RMSEA	0.047	<0.06	Excellent
PClose	0.925	>0.05	Excellent

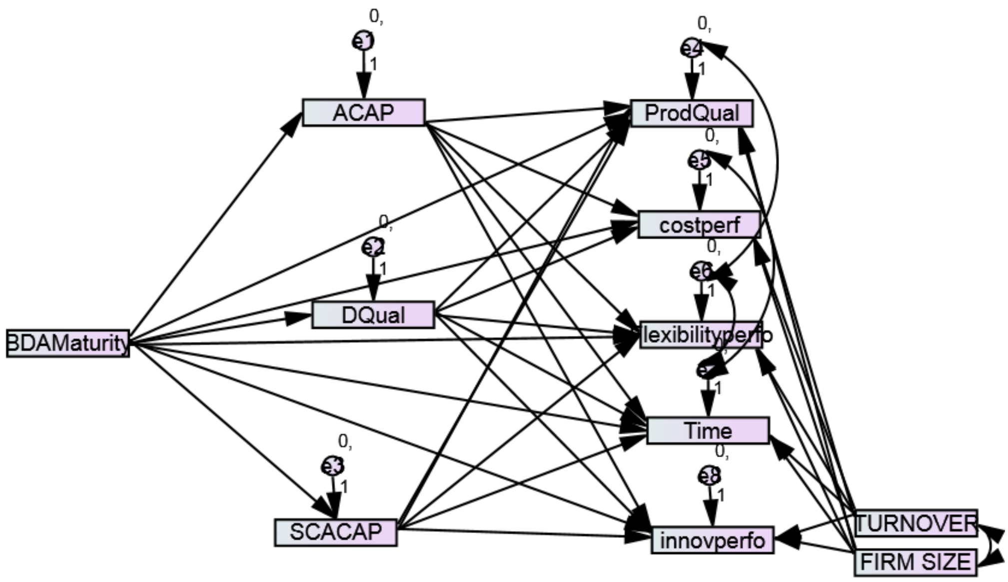


Figure 5.8 Structural model

5.5.2 Results of direct effects of BDA Maturity on ACAP, DIQ, SCA capability on firm performance dimensions.

Based on the theoretical model, it is hypothesised that the BDA maturity will have a significant positive effect on operational and innovation performance (H1, H2). As shown in Table 5.16, this research has found evidence supporting the hypothesis related to the direct impact of BDA maturity and operational and innovation performance. The direct path from BDA maturity →Product quality (0.595**), BDA maturity→Cost (0.615***), BDA Maturity →Flexibility (0.554***) and BDA maturity →Time (0.610***), are all found to be significant at the level of p<0.001. Moreover, the impact of BDA maturity on overall operational performance, created as a second-order reflective construct composed of product quality, cost, flexibility and time, is also tested (BDA→Overall operational performance) and the findings suggests there is a positive relationship between them. With regards to the impact of BDA on innovation (BDA maturity→innovation), the findings suggest (0.599***) there is a significant positive relationship.

Table 5.16 Direct effect of Second-order BDA maturity on Firm performance dimensions

	Un Std Estimate	**Std Estimates**	**S.E.**	**C.R. (or) Z statistics**	**P**	**Bootstrapping 5000 samples (95% Confident interval)**		
						Lower	Upper	P
BDA Maturity →Product Quality	0.601	.595 ***	0.078	7.714	***	0.477	0.686	0.001
BDA Maturity →Cost	0.642	.615 ***	0.085	7.564	***	0.495	0.716	0
BDA Maturity →Flexibility	0.612	.554 ***	0.086	7.114	***	0.421	0.666	0
BDA Maturity →Time	0.725	.610 ***	0.088	8.199	***	0.496	0.699	0.001
BDA Maturity →Innovation	0.669	.599 ***	0.086	7.742	***	0.462	0.704	0.001
BDA Maturity →overall operational performance	0.733	0.569***	0.069	6.681	***	0.418	0.727	0

Significance of Correlations: *** p < 0.001, ** p < 0.010, * p < 0.050, † p < 0.100

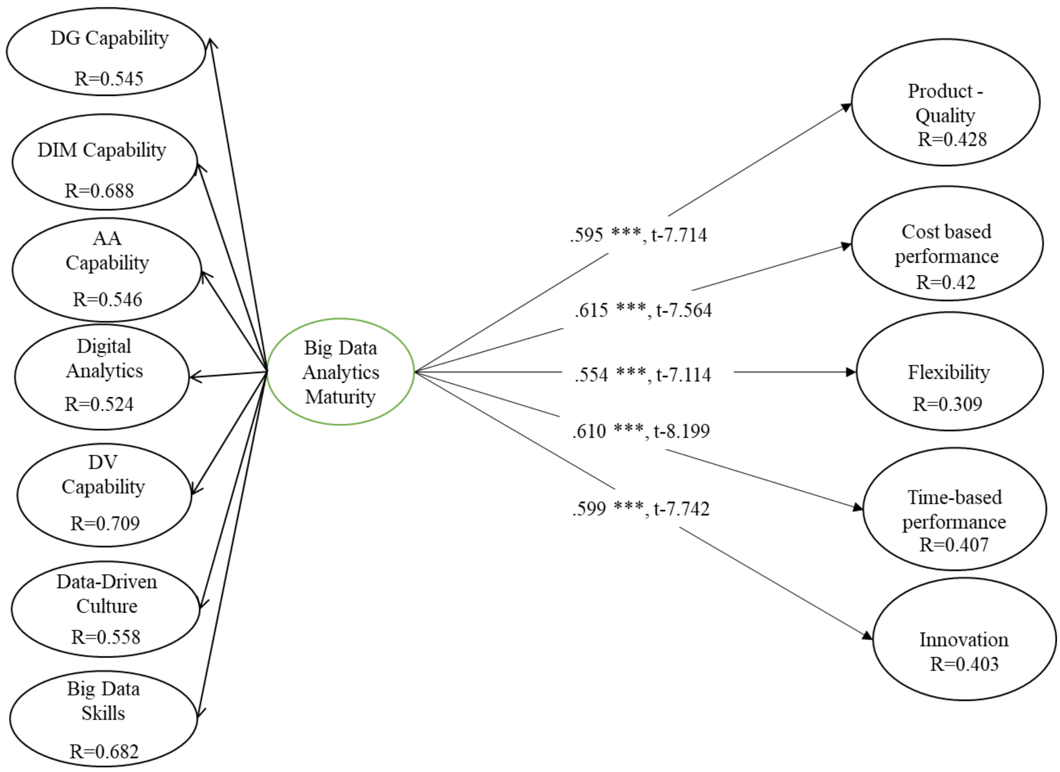


Figure 5.9 Results of the direct effect of BDA maturity on firm performance

Further, hypotheses H3, H4, and H5 state that BDA maturity positively influences the mediating variables absorptive capacity (BDA maturity→ACAP), data and information quality (BDA maturity →DIQ), and SCA capability (BDA maturity→SCA). These hypotheses are developed based on the concept of a hierarchic of dynamic capabilities (Liu et al. 2013). As given in Table 5.17, the findings confirm the positive influence of BDA maturity on all the three mediating variables.

Table 5.17 Results of the direct effect of the IV on MV and MV on DV

Predictor	Outcome	Std Beta
ACAP	Overall Operational Performance	0.44***
ACAP	Innovation Performance	0.399***
DIQ	Overall Operational Performance	0.331***
DIQ	Innovation Performance	0.179 (NS)
SCA capability	Overall Operational Performance	0.508***
SCA capability	Innovation Performance	0.216 *
BDA Maturity	ACAP	0.684***

BDA Maturity	Data and Information Quality	0.688***
BDA Maturity	SCA Capability	0.673 ***
Significance of Correlations: *** p < 0.001, ** p < 0.010, * p < 0.050, † p < 0.100		

Moreover, the effect of individual dimensions of BDA maturity on the operational and innovation performance is also examined. As given in Table 5.18, it is hypothesised that the seven dimensions of BDA maturity namely data generation, data integration, advanced analytics, digital analytics, data visualisation, big data skills, and data-driven culture are positively related to operational performance (H1e—H1k) and innovation performance (H2a-H2g). The testing of the following hypotheses, H1e (data generation capability →Overall operational performance), H1g (Advanced analytics capability →Overall operational performance), H1j (Big data skills →Overall operational performance) and H1k (data-driven culture →Overall operational performance) has confirmed that the key capabilities that influence operational performance are data generation, advanced analytics, big data skills and data-driven culture. However, in contrast, this research did not find sufficient evidence to support the claims proposed in H1f, H1h and H1k, as the results are non-significant and the effect sizes are also small. Similarly, H2a-H2g are tested to find the influence of BDA capabilities on the innovation performance of the UK manufacturing sector. Findings suggest that only data generation, digital analytics and data-driven culture have a significant positive relationship with innovation performance, supporting H2a, H2d, H2g. However, there is no evidence to support H2b, H2c, H2e and H2f. This reveals that the three most important capabilities to enhance innovation are data generation, digital analytics and data-driven culture. Data-driven culture especially is found to have a significantly greater influence on improving innovation performance, as is evident from the effect size and significance level (0.525***). This study did not find enough evidence to support the claim of the positive influence of some of the BDA capabilities such as data integration, data visualisation on neither operational performance nor innovation performance. However, the findings have provided original insights about the key capabilities that influence value creation. Moreover, since there is sufficient evidence regarding the positive relationship between BDA maturity and firm performance dimensions, it can be argued that organisations should holistically adopt all the capabilities discussed in this research to utilise the full potential of BDA. A detailed explanation of these findings from direct effect analysis is provided in the discussions chapter 6.

Table 5.18 Direct effect of First-order BDA capabilities on Firm performance dimensions

Predictor	Outcome	Std Beta
Data generation	Overall Operational Perf	.284 **
Data integration	Overall Operational Perf	-0.077 (NS)
Advanced analytics	Overall Operational Perf	.247 **
Digital analytics	Overall Operational Perf	.143 †
Data Visualisation	Overall Operational Perf	-0.089 (NS)
Big Data skills	Overall Operational Perf	.237 *
Data culture	Overall Operational Perf	.194 *
Data generation	Innovation Performance	.273 **
Data integration	Innovation Performance	-0.173 (NS)
Advanced analytics	Innovation Performance	0.109 (NS)
Digital analytics	Innovation Performance	.246 **
Data Visualisation	Innovation Performance	-0.146 (NS)
Big Data skills	Innovation Performance	-0.104 (NS)
Data Driven culture	Innovation Performance	.525 ***

5.5.3 Mediating effect of ACAP on the relationship between BDA Maturity and individual firm performance dimensions

Besides the direct effect of BDA maturity on operational and innovation performance, this study investigated the mediating role of absorptive capacity (ACAP) on the established relationship between BDA maturity and firm performance dimensions. As discussed earlier, bootstrapping and the product-of-coefficient approach are used to test the role of the intervening variable. For testing mediation in AMOS, user-defined estimates are created to calculate product coefficients of the path 'a' and 'b' and the significance level of the indirect effect is estimated via bootstrapping and a 95% bias-corrected confidence interval. Hypotheses H7a-d, H8 state that absorptive capacity as a dynamic capability mediates the positive effect of BDA maturity on individual operational performances and innovation performance. Table 5.19, provides the results of the mediation analysis. Findings suggest that ACAP significantly mediate the positive relationship, supporting H7a-d and H8. However, although the effect size of the direct effect of BDA maturity on operation and innovation performance is reduced when adjusted for the mediating variable, the effect is still significant. Hence, the type of mediation by ACAP on the relationship is complementary in nature (Zhao et al. 2010). A detailed discussion of these findings is provided in chapter 6.

Table 5.19 Results on the mediating role of absorptive capacity

Hypothesis	Direct effect without mediation (Standardised estimates)	The direct effect with mediation (Mediator =Absorptive capacity)	Indirect effect	Bootstrap (5000 samples) 95% Confident interval			Remarks
				Lower	Upper	P	
BDA Maturity → ACAP → Product Quality	.595 ***	0.289***	0.269***	0.131	0.427	0	Complementary Mediation
BDA Maturity → ACAP → Cost	.615 ***	0.336***	0.241**	0.089	0.424	0.003	Complementary Mediation
BDA Maturity → ACAP → Flexibility	.554 ***	0.347***	0.187*	0.007	0.365	0.043	Complementary Mediation
BDA Maturity → ACAP → Time	.610 ***	0.378***	0.221**	0.042	0.418	0.018	Complementary Mediation
BDA Maturity → ACAP → Innovation	.599 ***	0.281***	0.304***	0.125	0.484	0	Complementary Mediation
BDA Maturity → ACAP → overall operational performance	0.569***	0.419***	0.301***	0.154	0.454	0	Complementary Mediation
Significance of Correlations: *** p < 0.001, ** p < 0.010, * p < 0.050, † p < 0.100							

Table 5.20 Estimation of R square

R Square	Estimate
Data and Information Quality	0.494
Absorptive capacity	0.518
SCA capability	0.548
Digital Analytics	0.524
Flexibility	0.309
DIM Capability	0.688
Innovation	0.403
Advanced Analytics	0.546
Data Visualisation	0.709
Product Quality	0.428
Cost-based performance	0.420
Data generation capability	0.545
Big Data skills	0.682
Data-driven culture	0.558
Time-based performance	0.407

5.5.4 Mediating effect of DIQ on the relationship between BDA Maturity and firm performance dimensions

The results of the mediation of data and information quality on the relationship between BDA maturity and firm performance dimensions are given in Table 5.21. H10a-d and H11 propose that data and information quality as a consequent of BDA maturity mediates its relationships with operational and innovation performance. Here, the intention is to test the importance of resource quality, i.e. data and information quality on value creation. Consequently, the findings have provided adequate evidence to support H10a (BDA Maturity → DIQ → Product Quality), H10b (BDA Maturity → DIQ → cost), and H10d (BDA Maturity → DIQ → time), suggesting complementary mediation. However, the relationship between BDA Maturity → flexibility and BDA Maturity → Innovation is not mediated by DIQ, as a 95% confidence interval shows a non-significant result. In these cases, it is concluded that the effect is direct-only-non-mediation. It could be argued that data and information quality is a technical consequent of BDA maturity, but innovation is largely behavioural in nature. The possible explanations for these contrasting results will be provided in the discussion chapter.

Table 5.21 Results on the mediating role of data and information quality

Hypothesis	Direct effect without mediation (Standardised estimates)	Direct effect with mediation (Mediator =DIQ)	Indirect effect	Bootstrap (5000 samples)			Remarks
				Lower	Upper	P	
BDA Maturity → DIQ → Product Quality	.595 ***	0.439***	0.137	0.002	0.306	0.045	Complementary Mediation
BDA Maturity → DIQ → Cost	.615 ***	0.408***	0.181	0.03	0.376	0.019	Complementary Mediation
BDA Maturity → DIQ → Flexibility	.554 ***	0.419***	0.121	-0.049	0.295	0.149	Direct-only Non-Mediation
BDA Maturity → DIQ → Time	.610 ***	0.4***	0.212	0.048	0.423	0.014	Complementary Mediation
BDA Maturity → DIQ → Innovation	.599 ***	0.45***	0.138	-0.026	0.321	0.092	Direct-only Non-Mediation
BDA Maturity → DIQ → overall operational performance	0.569***	0.484***	0.227	0.103	0.381	0.001	Complementary Mediation

Significance of Correlations: *** p < 0.001, ** p < 0.010, * p < 0.050, † p < 0.100

5.5.5 Mediating effect of SCA capability on the relationship between BDA Maturity and firm performance dimensions

Similar to BDA capabilities, Supply Chain Analytics (SCA) capabilities is also conceptualised as a dynamic capability. While the BDA maturity construct measures the capabilities possessed by the organisation at the intra-organisational level, the potential role of application of BDA at the supply chain level is not explained in previous research. Thus, in this research, it is proposed in hypotheses H13, H13a-d, H14 that the practice of Big data and analytics at supply chain level would mediate the positive effect of BDA maturity (organisation level) on the operational and innovation performance. The results shown in Table 5.22 indicate that the SCA capability mediates the positive relationship between BDA maturity and operational performance dimensions such as product quality, cost, flexibility and time, thus supporting H13a-d. However, its mediating effect on the path BDA maturity → innovation is found to be non-significant as indicated by the 95% confidence interval estimated with 5,000 bootstrap sample. Hence, only hypotheses H13a-d are accepted, showing complementary mediation. However, H14 (BDA Maturity → SCA→Innovation) is not supported, indicating direct-only non-mediation results, as there is not enough evidence to support the claim. The implications of these findings for BDA practice will be further explored in the discussion chapter.

Table 5.22 Results on the mediating role of SCA capability'

Hypothesis	Direct effect without mediation (Std estimates)	Direct effect with mediation (Mediator =SCA)	Indirect effect	Bootstrap (5000 samples)			Remarks
				Lower	Upper	P	
BDA Maturity → SCA → Product Quality	.595 ***	0.208*	0.361	0.178	0.603	0	Complementary Mediation
BDA Maturity → SCA → Cost	.615 ***	0.263*	0.314	0.113	0.61	0.003	Complementary Mediation
BDA Maturity → SCA → Flexibility	.554 ***	0.272*	0.25	0.059	0.567	0.015	Complementary Mediation
BDA Maturity → SCA → Time	.610 ***	0.276**	0.292	0.11	0.68	0.007	Complementary Mediation
BDA Maturity → SCA → Innovation	.599 ***	0.371***	0.198	-0.042	0.476	0.092	Direct-only Non-Mediation
BDA Maturity -->SCA-->overall operational performance	0.569***	0.338**	0.394***	0.234	0.578	0	Complementary Mediation
Significance of Correlations: *** p < 0.001, ** p < 0.010, * p < 0.050, † p < 0.100							

Finally, as shown in the structural model (Figure 5.8), both the direct effect model and the multiple mediation models are controlled. The two control variables used in the research are a number of employees and annual turnover. It is assumed that large firms may achieve a higher level of effect on performance compared to SMEs. However, as given in Table 5.23, it is found out that the controls have no effect on the endogenous variables.

Table 5.23 Effect of control variables

Predictor	Outcome	Std Beta
Firm Size (no of employees)	Overall Operational Perf	-0.101 (NS)
Turnover	Overall Operational Perf	0.082 (NS)
Firm Size (no of employees)	Innovation Performance	0.152 (NS)
Turnover	Innovation Performance	-0.012 (NS)

5.5.6 Summary of explanatory data analysis

In summary, the findings presented in section 5.5 validated the hypothesis proposed in chapter 3 (theory). Table 5.24 contains the list of hypotheses tested using Structural Equation Modelling (SEM) technique. In the case of direct effects, the findings confirmed the positive relationship between BDA maturity and firm performance dimensions. Moreover, the key capabilities that contribute to enhancing performance are also identified. Further, findings confirmed that ACAP is a partial mediator on the relationship between BDA maturity and operational and innovation performance. However, the intervening variable 'data and information quality' and 'SCA capability' only mediated the path between BDA maturity and operational performance, but its role on the relationship between BDA maturity and innovation is not significant based on the findings of this research. In the next chapter, the findings and its inferences will be elaborated along with the contributions of this research to the theory and practice of BDA in the UK manufacturing sector.

Table 5.24 Results of hypothesis testing

No	Hypothesis	Results
RQ1a.	**What is the relationship between Big Data Analytics capability maturity and firm performance dimensions?**	
H1	Big data analytics maturity has a significant positive effect on operational performance dimensions	Supported
	H1a: Big data analytics maturity has a significant positive effect on product quality based operational performance	Supported
	H1b: Big data analytics maturity has a significant positive effect on cost-based operational performance	Supported

	H1c: Big data analytics maturity has a significant positive effect on flexibility based of operational performance	Supported
	H1d: Big data analytics maturity has a significant positive effect on time-based operational performance	Supported
	H1e: Data Generation capability has a significant positive effect on overall operational performance	
	H1f: Data Integration capability has a significant positive effect on overall operational performance	Not Supported
	H1g: Advanced Analytics capability has a significant positive effect on overall operational performance	Supported
	H1h: Digital Analytics capability has a significant positive effect on overall operational performance	Not Supported
	H1i: Data Visualisation capability has a significant positive effect on overall operational performance	Not Supported
	H1j: Big Data Skills has a significant positive effect on overall operational performance	Supported
	H1k: Data-Driven Culture has a significant positive effect on overall operational performance	Supported
H2	Big data analytics maturity has a significant positive effect on innovation performance.	Supported
	H2a: Data Generation capability has a significant positive effect on Innovation performance.	Supported
	H2b: Data Integration capability has a significant positive effect on Innovation performance.	Not Supported
	H2c: Advanced Analytics capability has a significant positive effect on Innovation performance.	Not Supported
	H2d: Digital Analytics capability has a significant positive effect on Innovation performance.	Supported
	H2e: Data Visualisation capability has a significant positive effect on Innovation performance.	Not Supported
	H2f: Big Data Skills has a significant positive effect on Innovation performance.	Not Supported
	H2g: Data-Driven Culture has a significant positive effect on Innovation performance.	Supported
H3	Big data analytics maturity has a significant positive effect on Absorptive capacity	Supported
H4	Big data analytics maturity has a significant positive effect on data and information quality	Supported
H5	Big data analytics maturity has a significant positive effect on supply chain analytics capability	Supported
RQ1b.	**What is the role of Absorptive Capacity on the relationship between BDA capability maturity and firm performance?**	
H6a	Absorptive capacity is positively related to operational performance.	Supported
H6b	Absorptive capacity is positively related to innovation performance.	Supported
H7	Absorptive capacity mediates the relationship between BDA maturity and operational performance.	Supported

	H7a: Absorptive capacity mediates the positive effect of Big Data Analytics maturity on product quality based operational performance	Supported
	H7b: Absorptive capacity mediates the positive effect of Big Data Analytics maturity on cost-based operational performance	Supported
	H7c: Absorptive capacity mediates the positive effect of Big Data Analytics maturity on flexibility based of operational performance	Supported
	H7d: Absorptive capacity mediates the positive effect of Big Data Analytics maturity on time-based operational performance	Supported
H8	Absorptive capacity will mediate the relationship between BDA maturity and innovation performance.	Supported
RQ1c.	**What is the role of Data and Information Quality (DIQ) on the relationship between BDA capability maturity and firm performance?**	
H9a	The operational performance of an organisation is positively affected by its ability to maintain data and information quality.	Supported
H9b	The innovation performance of an organisation is positively affected by its ability to maintain data and information quality.	Not Supported
H10	DIQ mediates the relationship between BDA maturity and operational performance.	Supported
	H10a: DIQ mediates the positive effect of Big Data Analytics maturity on product quality based operational performance.	Supported
	H10b: DIQ mediates the positive effect of Big Data Analytics maturity on cost-based operational performance	Supported
	H10c: DIQ mediates the positive effect of Big Data Analytics maturity on flexibility based of operational performance	Not Supported
	H10d: DIQ mediates the positive effect of Big Data Analytics maturity on time-based operational performance	Supported
H11	DIQ will mediate the relationship between BDA maturity and innovation performance.	Not Supported
RQ1d.	**What is the role of Supply Chain Analytics capability (SCA) on the relationship between BDA capability maturity and firm performance**	
H12a	Supply chain analytics capabilities is positively related to operational performance.	Supported
H12b	Supply chain analytics capabilities is positively related to innovation performance.	Supported
H13	Supply chain analytics capabilities mediates the relationship between BDA maturity and operational performance.	Supported
	H13a: SCA mediates the positive effect of Big Data Analytics maturity on product quality based operational performance.	Supported
	H13b: SCA mediates the positive effect of Big Data Analytics maturity on cost-based operational performance.	Supported
	H13c: SCA mediates the positive effect of Big Data Analytics maturity on flexibility based of operational performance.	Supported
	H13d: SCA mediates the positive effect of Big Data Analytics maturity on time-based operational performance.	Supported
H14	Supply chain analytics capabilities will mediate the relationship between BDA maturity and innovation performance.	Not Supported

5.6 Cluster Analysis

This section elaborates the implementation and findings of cluster analysis used in this research to explore the phenomenon of the digital divide caused by the disparity in the adoption trend of BDA in the UK's manufacturing industry. The clustering analysis is performed based on the steps discussed in Xu and Wunsch (2005). The statistical description of data (provided in Appendix D) contains the list of variables in the data set and the descriptive statistics. Appropriate distance measures are considered in the next section before the implementation of the clustering algorithms. Since the clustering technique is exploratory in nature, experiments are conducted with varying number of clusters using three different algorithms and the results are validated statistically. The three clustering algorithms implemented in this research are Hierarchical clustering, K-medoids (PAM) and Latent Class Analysis (LCA). Hierarchical clustering and K-medoids (PAM) are traditional clustering techniques, but LCA is a model-based technique used to identify homogeneous groups in the data. The foundation and description of these techniques are provided in the methodology chapter 4.

5.6.1 Distance measures

The distance between two objects or observations is generally used to measure the dissimilarity or similarity between them. There are many ways to calculate the distance between observations using metrics such as Minkowski distance and Mahalanobis distance, but the most popular one widely used in the literature and practice is the 'Euclidean distance' (Xu and Wunsch, 2005). The expression to calculate Euclidean distance 'd' of two observations X and Y is given below.

$$\boldsymbol{d} = \sqrt{\sum}(\boldsymbol{x} - \boldsymbol{y})^{\boldsymbol{2}} \text{ (Euclidean distance equation)} \qquad (1)$$

Whereas, 'Manhattan' distance calculates the sum of absolute differences.

$$d(i,j) = |X_{i1} - X_{j1}| + |X_{i2} - X_{j2}| + \cdots + |X_{ip} - X_{jp}| \qquad (2)$$

(Manhattan distance equation**)**

However, since the data is ordinal in this study, it could be argued that the distance between the two categories would not be the same. In this research, numerical value: '1' represents 'Strongly disagree', '2' represents 'disagree', '3' represents 'Neither agree nor disagree', '4' represents 'agree, and '5' represents 'strongly agree', and each level has a particular rank. Choosing a right distance metric is critical for the successful implementation of a clustering algorithm and it should be truly based on the data type. R has two relevant packages ‘cluster’ and ‘clusterSim’, which is used to measure the distance of ordinal data types. The package ‘cluster’ has a function called ‘daisy’ which calculates the ‘general dissimilarity coefficient of Gower’, suitable for mixed and ordinal data types.

$$d(i,j) = \frac{1}{p}\sum_{i=1}^{p} d_{ij}^{(f)} \text{ (Gower distance equation)} \qquad (3)$$

According to Gower (1971), in order to calculate the Gower dissimilarity matrix, the variables are standardised and the distance between two vectors is measured based on “the sum of all the variable-specific distances”. With the Gower metric, each variable is standardised by dividing vectors with the range of a particular variable and subtracting it with the minimum value, and the final scale of variables will have values in the range (0, 1). Similarly, ‘ClusterSim’ package also has a function called ‘GDM2’ (Generalized Distance Measure) which is argued to be more suitable to variables with ordinal data types. Walesiak (1999) and Jajuga et al. (2003) have described the method of the generalised distance metric (GDM) to measure dissimilarity between observations as it is based on the concept of generalised correlation coefficient.

$$d_{ik} = \frac{1-S_{ik}}{2} = \frac{1}{2} - \frac{\sum_{j=1}^{m} a_{ikj}\, b_{kij} + \sum_{j=1}^{m}\sum_{\substack{l=1 \\ l\neq i,k}}^{n} a_{ilj}\, b_{klj}}{\left[\sum_{j=1}^{m}\sum_{l=1}^{n} a_{ilj}^{2} \sum_{j=1}^{m}\sum_{l=1}^{n} b_{klj}^{2}\right]^{\frac{1}{2}}} \qquad (4)$$

Where,

$d_{ik}(s_{ik})$ – proximity measure,

i, k, l – indicates number of objects 1 to n,

j – indicates the number of variables 1 to m,

Also, for an ordinal data scale ‘a_{ipj}’ and ‘bkrj’ in the above equation is given as

$$a_{ipj}(b_{krj}) = \begin{cases} 1 \quad if\ x_{ij} > x_{pj}(x_{kj} > x_{rj}) \\ 0 \quad if\ x_{ij} = x_{pj}(x_{kj} = x_{rj}) \\ -1 \quad if\ x_{ij} < x_{pj}(x_{kj} < x_{rj}) \end{cases} \quad for\ p = k, l; r = i, l.$$

(5)

Nevertheless, Xu and Wunsch (2005) argued that the choice of distance metrics is often subjective and based on the ability to generate interesting clusters. In this research, four different distance metrics, Euclidean distance, Manhattan distance, Gower's metric, and GDM distance measures, are used to convert the raw data into the dissimilarity matrix and treated as an input to the clustering algorithms. However, model-based clustering techniques such as 'Latent class analysis' can take the raw data which is ordinal in nature and do not require any prior transformation.

In this research, the four distance matrix are calculated in R using the two libraries 'Cluster', and 'ClusterSim' which contains 'daisy' and 'GDM2' functions respectively to evaluate these distance measures.

5.6.2 Hierarchical clustering

As described in (Arunachalam and Kumar 2018), Hierarchical cluster analysis (HCA) can be done using two different methods Bottom-up (Agglomerative) or Top-down (Divisive). From the literature review, it is evident that majority of market segmentation studies have used Agglomerative clustering technique especially the Ward method. In Agglomerative clustering, each observation is considered as its own cluster and it is joined with a neighbouring cluster based on the similarity between their distances, and the process repeats until all the observations are connected. The dissimilarity matrix calculated using the four-distance metrics discussed above is used as the input for Hierarchical clustering.

The implementation of hierarchical clustering in R studio is performed using the 'cluster' package (Kaufman and Rousseeuw 1990). There are several hierarchical techniques based on the linkage method such as Ward (Murtagh and Legendre 2014) and 'Complete'. The cophenetic correlation coefficient is calculated, which is the correlation between the distance matrix and the cophenetic matrix. A cophenetic correlation coefficient (c) value close to 1 indicates that the clustering is a good fit with the nature of the data (Saraçli et al. 2013). From Table 5.25, it is evident that in most cases 'Ward' and 'average' linkage methods of hierarchical clustering are found to be a good fit for the data. Consequently, several experiments are conducted using these two hierarchical linkage methods and the results are given in Table 5.26. The 2, 3, 4, and 5 cluster solutions are obtained using 'hclust' function in R and the results are validated. Cluster validity is measured in R with the help of a specific package called 'CValid', which has an inbuilt function to validate the clustering solutions. From Table 5.26, it can be seen that both the Silhouette index and the DB index have revealed either 3 or 4 cluster solutions would be optimum for the dataset. It should also be noted that the low silhouette value (< 0.5) indicates the possibility of an artificial partitioning of data. In our experiments, several clustering solutions have resulted in silhouette values close to zero and therefore it can be argued that there are different objects that overlap and lie between the clusters (Kaufman and Rousseeuw 1990). Kaufman and Rousseeuw (1990) suggested that only a silhouette value between 0.51 and 0.70 indicates the formation of clusters with a reasonable structure. Therefore, clustering solutions which meet this threshold silhouette index criterion are considered significant in this study. However, with respect to linkage methods, only the 'Ward' method has relatively a high silhouette value. Moreover, low values of Xie-Beni and DB index indicate a good clustering solution, but in the case of the silhouette and Dunn Index high values are desirable. Silhouette, DB and Xie-Beni index calculated for 'Ward' based clustering indicate 3 clusters as optimum, but also reveals the possibility of 4 clusters. Similar to the findings of (Arunachalam and Kumar 2018), the output of the average linkage method is also validated and suggests 4 or 5 clusters, but its dendrograms are messy and irregular in shape and it is difficult to visually examine the number of clusters present in the data set.

Table 5.25 Hierarchical clustering experiments and cophenetic coefficient value.

Exp No	Hierarchical clustering Experiments and Cophenetic value	
1	**Euclidean distance with ward method**	**0.64064**
2	**Euclidean distance with complete method**	**0.6514595**
3	Euclidean distance with single method	0.3581455
4	Euclidean distance with centroid method	0.5978201
5	**Euclidean distance with average method**	**0.7100185**
6	**Manhattan distance with ward method**	**0.6412214**
7	Manhattan distance with complete method	0.5665678
8	Manhattan distance with single method	0.3005792
9	Manhattan distance with centroid method	0.5936742
10	**Manhattan distance with average method**	**0.6967792**
11	**Gower distance with ward method**	**0.6394638**
12	Gower distance with complete method	0.6159108
13	Gower distance with single method	0.2791569
14	Gower distance with centroid method	0.6082733
15	**Gower distance with average method**	**0.6891562**
16	**GDM with ward method**	**0.6205567**
17	GDM with complete method	0.5796027
18	GDM with single method	0.2205317
19	GDM with centroid method	0.6188396
20	**GDM with average method**	**0.667417**

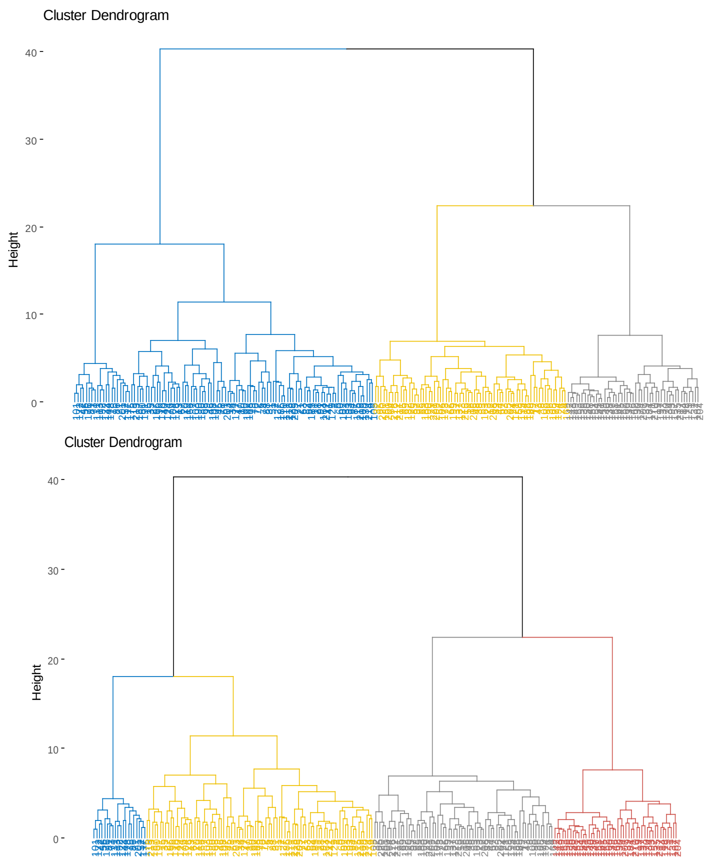


Figure 5.10 Dendrograms of 3 and 4 cluster solution -Ward method and Gower's distance matrix

Table 5.26 Hierarchical clustering experiments and cluster validations.

Hierarchical-Kmeans Clustering		**Ward method (ward.D2)**				**Average**			
No of clusters		2	3	4	5	2	3	4	5
Silhouette	GDM	0.307201	0.2982172	0.4775539	0.43203227	0.4056126	0.43453159	0.42276031	0.42032795
	Gower's	0.4397251	0.383427	0.4613013	0.43217063	0.4006028	0.4383427	NaN	NaN
	Manhattan	0.4397251	0.4908353	0.42634796	0.4318817	0.410082	0.43908353	NaN	NaN
	Euclidean	0.4397251	**0.5301948**	**0.52103562**	0.46051626	0.4005988	0.4383426	NaN	NaN
DB Index	GDM	1.159124	1.120953	0.8967987	1.800255	0.9165336	0.9060642	0.9529265	1.598968
	Gower's	0.8506573	0.8443719	0.9568228	1.001397	0.9312779	0.8443719	**0.4202925**	**0.4207229**
	Manhattan	0.8506573	**0.8239395**	0.911584	1.138142	0.9130843	0.8239395	**0.4175065**	0.4335586
	Euclidean	0.8506573	1.022652	1.305542	1.064828	0.9324959	0.8452788	0.5535201	1.002979
Dunn Index	GDM	0.1518187	0.1058972	0.1221599	0.1221599	0.1461625	0.1830028	0.1945264	0.1945264
	Gower's	0.147263	0.1987748	**0.2057567**	**0.2111361**	0.1750925	0.1987748	0.1987748	**0.2081521**
	Manhattan	0.147263	0.1987748	0.1868364	0.1694768	0.1856143	0.1987748	0.1987748	0.1868364
	Euclidean	0.147263	0.2148972	0.1723689	0.1789357	0.1750925	0.1987748	0.1987748	0.1987748
Xie-Beni Index	GDM	4.215796	7.179338	5.577843	5.248614	4.631547	3.277334	2.918838	2.837654
	Gower's	4.011435	**2.823893**	**2.990477**	**2.833598**	2.909826	2.823893	2.784888	**2.660364**
	Manhattan	4.011435	2.857159	3.081087	3.527933	2.635651	2.857159	2.81771	2.98695
	Euclidean	4.011435	2.728684	3.874779	3.534549	2.928481	2.837985	2.805129	2.755455
Note:	* Small Value of DB index & XB index indicates Compact and separate clustering and therefore minimised *Silhoutte & Dunn should be maximised								

5.6.3 K-Medoids clustering

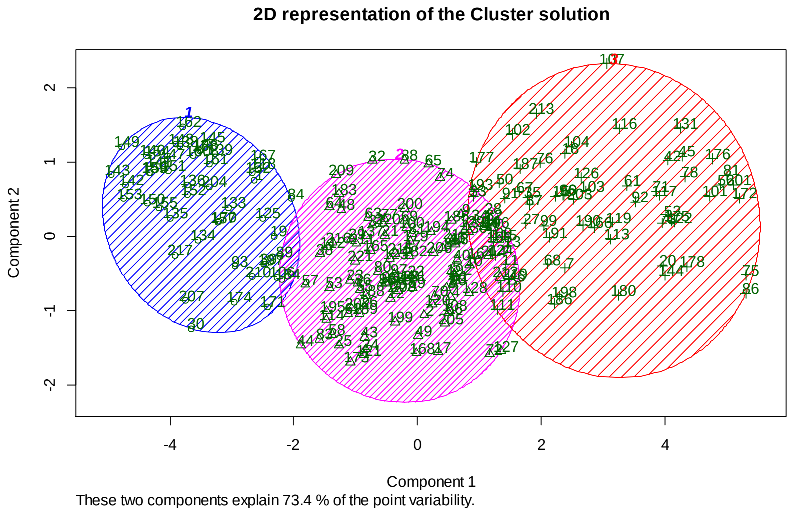
The K-means clustering algorithm for identifying homogenous groups is often used in practice (Dolničar 2003), but it can be argued that it is mainly suitable for interval data types. However, there are several versions of partitioning-based algorithms developed to overcome the drawbacks of K-means, and K-medoids is generally considered as being more robust and suitable for ordinal data sets, as clustering is done based on medoids unlike the K-means algorithm.

One of the popular k-Medoids algorithms, PAM (Partitioning around medoids), introduced by Kauffman and Rousseeuw (1990), is adopted in this research. According to Kauffman and Rousseeuw (1990), "k-Medoids minimizes a sum of dissimilarities instead of a sum of squared Euclidean distances". The procedure discussed in (Arunachalam and Kumar 2018) is used to implement PAM algorithm, and there are two important parameters to be considered; the distance metric 'd' and number of clusters 'k'. The justification for the choices of distance metrics has already been discussed in previous sections. The input argument used in the algorithm can either be a raw data frame, data matrix, or a dissimilarity matrix. If the data frame is used as input, limited option of distance metrics is available, only 'Euclidean' and 'Manhattan distance' can be calculated as an inbuilt option of PAM algorithm in R. However, the PAM algorithm permits the use of a dissimilarity matrix calculated using 'dist' or 'daisy' functions in R. Accordingly, dissimilarity metrics are calculated using external functions in R like 'daisy', 'gower.dist' and 'GDM2'. Then, the algorithm randomly computes 'k' objects of medoid, which itself is an object of the cluster having minimal average dissimilarity to all the objects. The objective function of the algorithm is to minimize the sum of dissimilarities between the 'k' medoids and the objects nearest to them.

5.6.3.1 PAM Experiments and results

Several experiments are conducted with varying the number of 'k' values and are visualised using 'Clusplot'. Cluster plot is a useful tool to visualise the structure, size, and the position of clusters in a 2-dimensional space. Simultaneously, the validity of the clustering is also measured using 4 internal cluster validity indexes such as Silhouette Co-efficient, DB Index, Dunn Index, and Xie Beni (XB) index. Each index used is different in its own way of measuring validity, but principally calculates how compact the clusters are and how much each is separated from the other clusters.

In this study, among the four validity indices, output of Silhouette, DB index, and Dunn are considered significant for PAM clustering, as XB index is argued to be effective to measure mainly fuzzy clusters. Table 5.27 outlines the cluster validity measures used, and it is obvious that the Gower's distance metric and k=4 appear to perform better in all the instances. Moreover, when the cluster plot generated by the experiment with Gower's distance, K=3 and 4, is examined (Figure 5.11), the presence of 3 distinct clusters is evident. However, it is noticeable that the cluster 3 is not stable and it divides further.



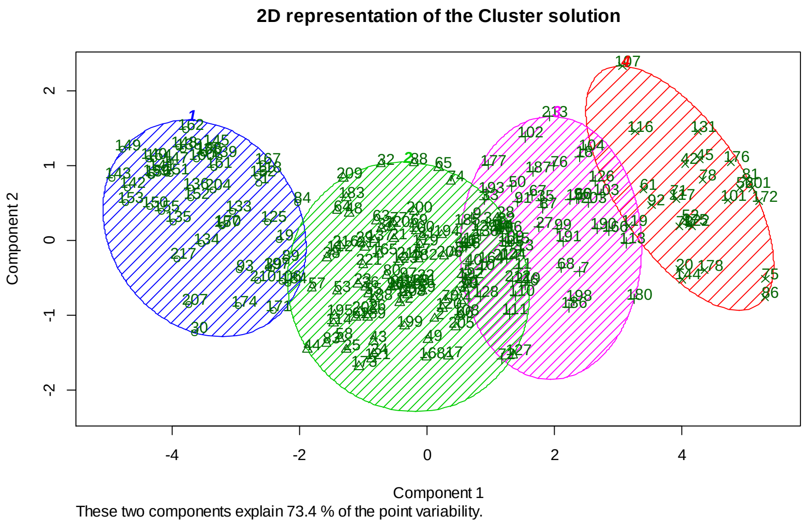


Figure 5.11 Cluster plot of PAM with Gower's 3 and 4 cluster solutions

Table 5.27 Cluster validity of PAM clustering experiments.

PAM experiments	Distance metric	Cluster number	Silhoutte Co-efficien	DB Index*	Dunn Index*	XB index
1	Gower's metrics	2	0.43239621	1.119441	0.161679	4.256627
	Generalised distance metrics	**2**	0.43443893	1.059465	0.1525409	4.823768
	Euclidean distance	2	**0.4914185**	0.9468776	0.147124	4.501138
	Manhattan distance	2	0.4320393	1.126287	0.1548609	3.942451
2	Gower's metrics	3	0.43083344	1.084094	0.1216549	8.078267
	Generalised distance metrics	3	0.42670089	1.173168	0.1108215	8.438667
	Euclidean distance	**3**	**0.53358354**	**0.914707**	0.1845773	3.226485
	Manhattan distance	3	0.4314325	1.063068	0.1319797	8.041819
3	Gower's metrics	**4**	**0.49718006**	0.9364028	**0.2067328**	**2.977169**
	Generalised distance metrics	4	0.43812553	1.045943	0.1712078	3.312605
	Euclidean distance	4	**0.5471811**	0.926596	0.1772462	3.684104
	Manhattan distance	4	0.42614065	0.9747385	0.1635524	4.517294
4	Gower's metrics	5	0.42151701	1.225917	0.1896159	3.388853
	Generalised distance metrics	**5**	0.41716695	1.003274	0.1297397	7.900647
	Euclidean distance	5	0.42131016	1.483965	0.1479087	5.084715
	Manhattan distance	5	0.42068881	1.225917	0.1635524	4.334065
Note:	* Small Value of DB index & XB index indicates Compact and separate clustering and therefore minimised					
	*Silhoutte & Dunn should be maximised					

Hence, PAM chooses the medoids randomly. From Table 5.27, it is observed that experiment 3a, in general, has achieved satisfactory results for all the four validity measures used (Silhouette -0.49718006, DB index-0.9364028, Dunn Index - 0.2067328, XB index – 2.977169). Varying results of PAM clustering for the same dataset is witnessed with respect to the different distance metrics used. PAM clustering with Euclidean distance has shown some significant advantage but it is mostly suitable for interval data types, which affects the quality of the clustering solution.

5.6.4 Latent Class analysis

The clustering algorithms implemented in the previous sections are either hierarchical or partition-based algorithms, and treat ordinal data as continuous. Moreover, in the case of other clustering approaches, the respondents are assigned to a specific group deterministically as the number of clusters is decided *apriori*. However, the LCA models assume the specific distribution of variables and assign the respondents to an unobserved latent class probabilistically (Todd 2013; Hagenaars and McCutcheon 2002). Further, it is argued that Latent Class Analysis (LCA) is more suitable for finding classes in ordinal data (Linzer and Lewis 2011). In this research, LCA is implemented in R using ‘PoLCA’ package, it identifies classes in an ordered categorical data. Since, the Likert scale is used to measure the dimensions of BDA, which is ordered from 1 –strongly disagree to 5- strongly agree, LCA can be considered more suitable to identify classes.

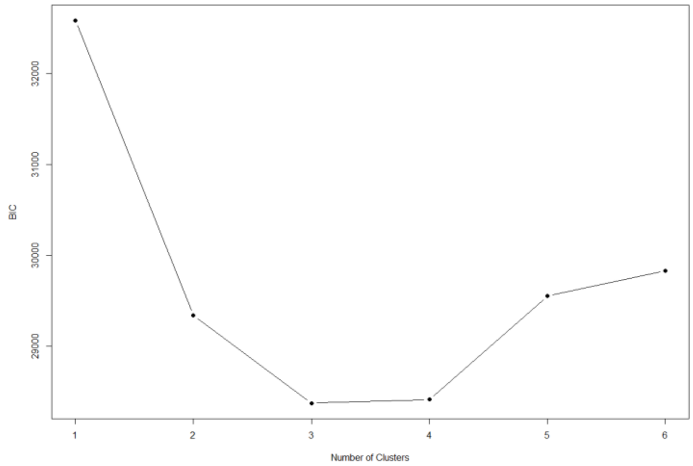


Figure 5.12 Bayesian Information Criterion for latent classes from 1 to 6

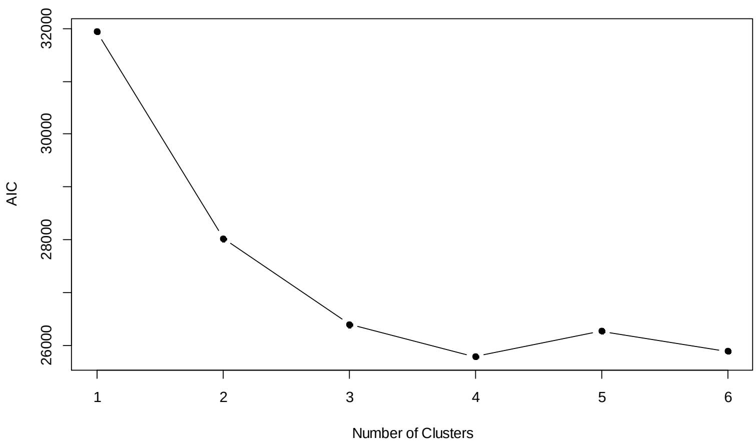


Figure 5.13 Akaike information criterion

In total, six different latent class models are tested and the results are given in Table 5.28. Todd (2013) have discussed the two information criteria that can be used to assess the goodness-of-fit statistics: 1. Bayesian Information Criterion (BIC), 2. Akaike Information Criterion (AIC). Results indicate that **(**Figure 5.12Figure 5.13**)**, both 3 and 4 class models have achieved low BIC value, and 4 class model received desirable low AIC value. In a 3-class latent model, respondents are categorised into 23.08%(51), 47.96 %(106) and 28.96 %(64), respectively. On the other hand, in case of 4 class latent model, the 221 responses are clustered into the proportion of 24.43 %(54), 16.28%(36), 23.53%(52), and 35.75%(79). However, the degrees of freedom (df) for most of the models have attained negative values. The degrees of freedom indicates the difference between the number of parameters estimated and the available observations. Since the sample size is 221 and the number of parameters estimated is more than the sample size the degrees of freedom value is negative and it could affect the statistical parameters estimated.

Table 5.28 Results of LCA models

	Modell	log likelihood	df	BIC	ABIC	CAIC	likelihood ratio
1	Model 1	-15774.5	29	32585.51	31977.05	32777.51	29165.8429
2	Model 2	-13631.2	-164	29340.75	28120.67	29725.75	24879.2431
3	Model 3	-12626.1	-357	28372.44	26540.73	28950.44	22869.08213
4	Model 4	-12125.5	-550	28412.91	25969.57	29183.91	21867.709
5	Model 5	-12173.7	-743	29551.32	26496.37	30515.32	21964.28018
6	Model 6	-11791.1	-936	29827.89	26161.3	30984.89	21198.99881
Note: Modell4- df is negative because the number of parameters estimated (771) exceeds the number of observations (221)							

5.6.5 Optimum cluster solution and cluster profiling

In summary, the clustering technique is exploratory in nature and so, to avoid potential bias while choosing the number of clusters, several experiments are conducted. The optimum number of clusters hidden in the data set is identified by assessing the clustering solutions using validity measures such as silhouette, Dunn index, and BIC. In most of the cases, the silhouette index suggests the presence of 3 clusters in the data set. In the Latent class analysis, the BIC criteria attained the desirable lowest score for the 3-class model. However, in contrast, the AIC criteria of LCA indicated the 4-class model. Moreover, Dunn and Xie-Beni validity indexes suggest the possibility of 4 clusters in the case of both Hierarchical clustering - ward linkage and K-medoids clustering. Hence, based on the clustering experiments, it is decided to continue with a 4 cluster solution for the profiling of clusters. Further, for practical reasons, the 4 cluster solution produced by K-medoids clustering with Gower's distance metrics is used for the profiling. The four clusters are cross profiled with two organisation specific demographic variables 'number of employees' and 'annual turnover'. Also, the other demographic variables measured in this research such as location and manufacturing sector produced an indistinct result not useful for descriptive purpose and hence excluded. For instance, the cross profile of 4 clusters and organisations' location is given in Figure 5.14. In most cases, the location information does not provide significant insights into the respective clusters and hence is disregarded to avoid potential biases with interpretation. Besides, the main purpose of conducting cluster analysis is to explore the digital divide between SMEs and large organisations in the UK. The cross-profiling of clusters with variables indicating firm size such as a number of employees and annual turnover is deemed to be sufficient. Subsequently, based on the characteristics, the four clusters are termed as 'Routine users', 'Ad-hoc users', 'Beginners/laggards', and 'Enthusiastic users'. The detailed discussion on the BDA maturity level of each clusters, its demographic characteristics and the phenomenon of the digital divide is provided in the discussions chapter 6.

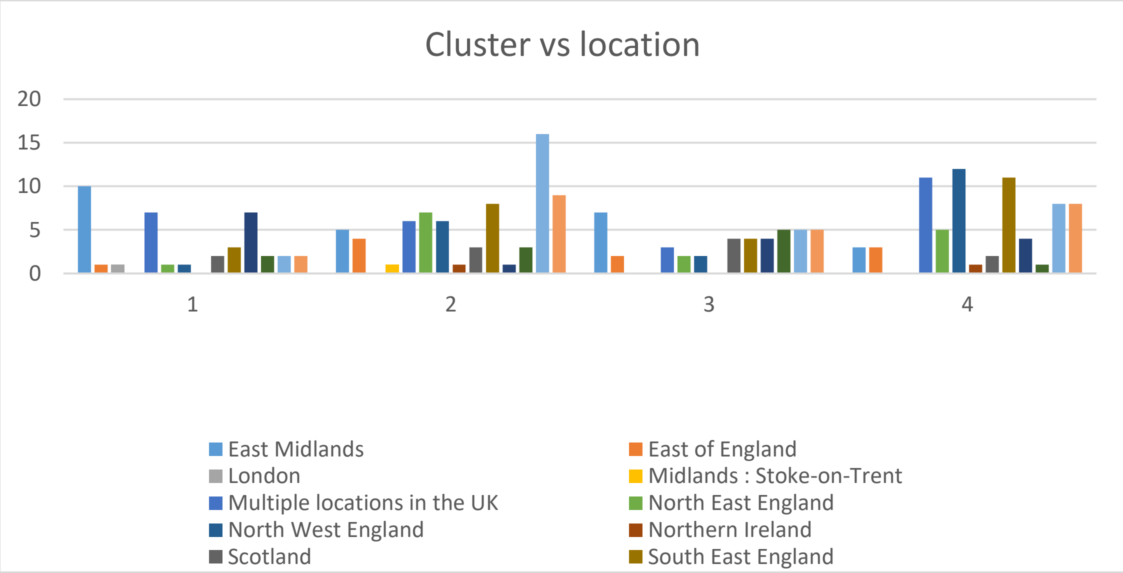


Figure 5.14 4 cross profile of the clusters vs Location

Table 5.29 cross profile of the clusters vs number of employees

		FIRM SIZE						Total
		1-9 employees	10-49 employees	50-99 employees	100-249 employees	250-500 employees	More than 500 employees	
Cluster Number of Case	1	2	3	1	9	12	12	39
	2	2	23	13	19	6	7	70
	3	5	17	2	10	5	4	43
	4	1	14	11	21	6	16	69
Total		10	57	27	59	29	39	221

Table 5.30 cross profile of the clusters vs annual turnover

		TURNOVER					Total
		<= 2Million	2 - 10 Million	10 - 25 million	25 - 50 million	> 50 million	
Cluster Number of Case	1	3	2	3	13	18	39
	2	12	22	16	10	10	70
	3	16	10	4	8	5	43
	4	4	24	16	8	17	69
Total		35	58	39	39	50	221

5.7 Summary of Data Analysis and findings

In this chapter, the research hypotheses proposed in chapter 3 are subjected to empirical investigation. Following the collection of data, the analysis is carried out as per the procedures discussed in the methodology chapter. The data is coded and pre-processed prior to the testing of the proposed relationship between the construct of the conceptual model. The findings of the descriptive analysis and factor analyses carried out in this study are also presented. Further, this chapter validated the proposed hypotheses. Findings show that ACAP, DIQ, and SCA partially mediate the relationship between BDA maturity and overall operational performance. However, with respect to the relationship between BDA maturity and innovation performance, it is not mediated by DIQ and SCA, but ACAP partially mediates the impact of BDA maturity on innovation performance. In addition, the use of cluster analysis revealed interesting insights into the phenomenon of the digital divide between firms in the UK manufacturing sector. In particular, this chapter illustrated that, in the UK manufacturing sector, there is a possibility of 4 clusters of organisations with the varying level of BDA maturity. The next chapter presents the discussions of these findings and highlights the significant contributions of the study to the theory, practice and policy of BDA practice.

Chapter 6 Discussions

6.1 Chapter Introduction

As discussed in chapter 1, the UK manufacturing sector is experiencing a dynamic uncertain environment influenced by various political and economic factors. This research has explored the potential of innovative BDA technology for providing adapting mechanisms to cope with the changing environment. Consequently, the aim of this research is twofold: first, to gain insights into the impact BDA capabilities have on the performance of the UK manufacturing firms; second, to investigate the phenomenon of the digital divide. By holistically applying RBV, DC view and the view of the hierarchy of capabilities, this research proposed a conceptual model in which BDA capabilities maturity, as lower-order capability, exert influence on operational and innovation performance through higher-order capabilities, such as ACAP, DIQ, and SCA capability. Using a quantitative approach, data were collected and analysed using structural equation modelling and cluster analysis (Chapter 5). This chapter discusses the results of the two empirical analysis performed in this research (Figure 6.1). In addition, the managerial implications and the contributions of this research are provided.

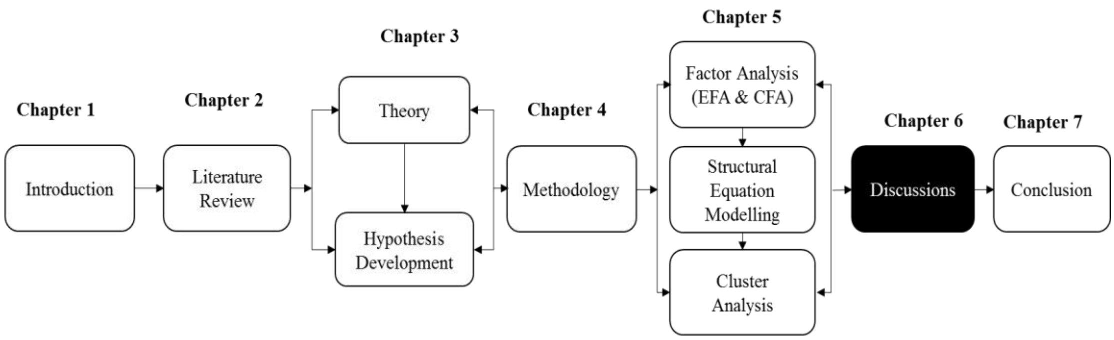


Figure 6.1 Position of the chapter in this thesis

6.2. Theoretical implications

Drawing upon the RBV, DCV, and hierarchy of capabilities view, this research has explored a relatively new domain of research on BDA practice and value creation. To the best of our knowledge, this is one of the first attempt to verify the proposed relationship with scientific rigour. This study provides a variety of implications and observations that stem from the research findings. This section begins with the discussions on theoretical implications based on the results of the data analysis and then elaborates on the managerial implications.

6.3 Explaining the impact of BDA capabilities on firm performance

6.3.1 Relationship between BDA and operational and innovation performance

Based on hypotheses (H1a-k) it is proposed that the seven dimensions of BDA capabilities (first-order) and the second-order BDA capabilities maturity significantly influence operational and innovation performance of the UK manufacturing firms. The findings from the analysis of the structural model provided evidence to support some of the proposed hypotheses. Figure 6.2, Figure 6.3 and Figure 6.4, illustrate the findings from the analysis of the direct effect of first-order BDA capabilities and second-order BDA capabilities maturity on operational and innovation performance. Consequently, it can be confirmed that BDA capabilities can improve the operational and innovation performance of manufacturing firms. The second-order construct BDA maturity is found to have a positive influence on operational performance dimensions such as product quality (0.595***), cost-based performance (0.615***), flexibility (0.554***) and time-based performance (0.610***). As well, it is also found to have a positive effect on the innovation performance (0.599***) of manufacturing firms (Figure 6.2).

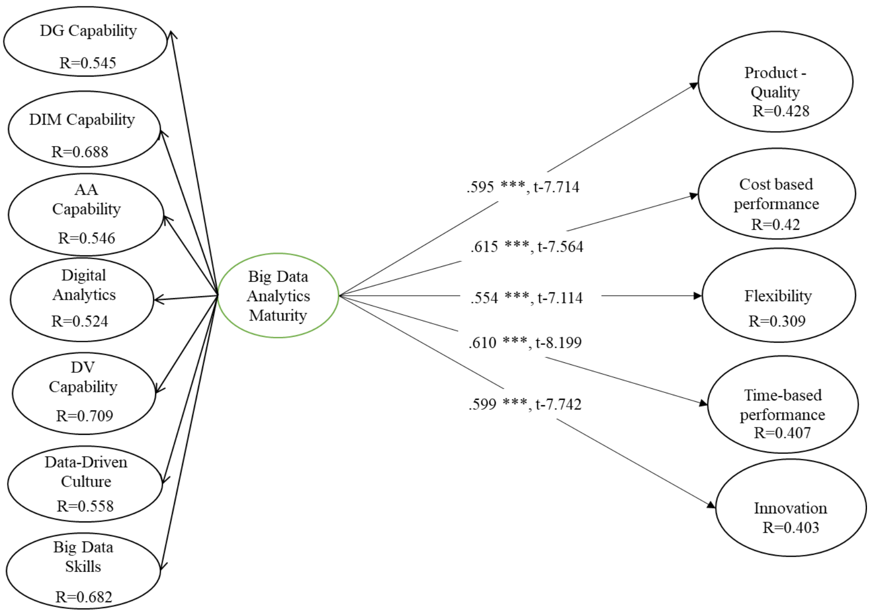


Figure 6.2 Relationship between Second-order BDA maturity and operational and performance

Moreover, the influence of first-order BDA capabilities revealed interesting insights into the capabilities that can positively influence performance dimensions and those that do not. In terms of improving operational performance, capabilities such as data generation (0.284**), advanced analytics (0.247**), digital analytics (0.143+), data-driven culture (0.194*) and big data skills (0.237*) are found to be significant (Figure 6.3). Whereas, for improving innovation performance of manufacturing firms, capabilities such as data generation (0.237**), advanced analytics (0.109*), digital analytics (0.246**), and data-driven culture (0.525***) are found to be more significant (Figure 6.4).

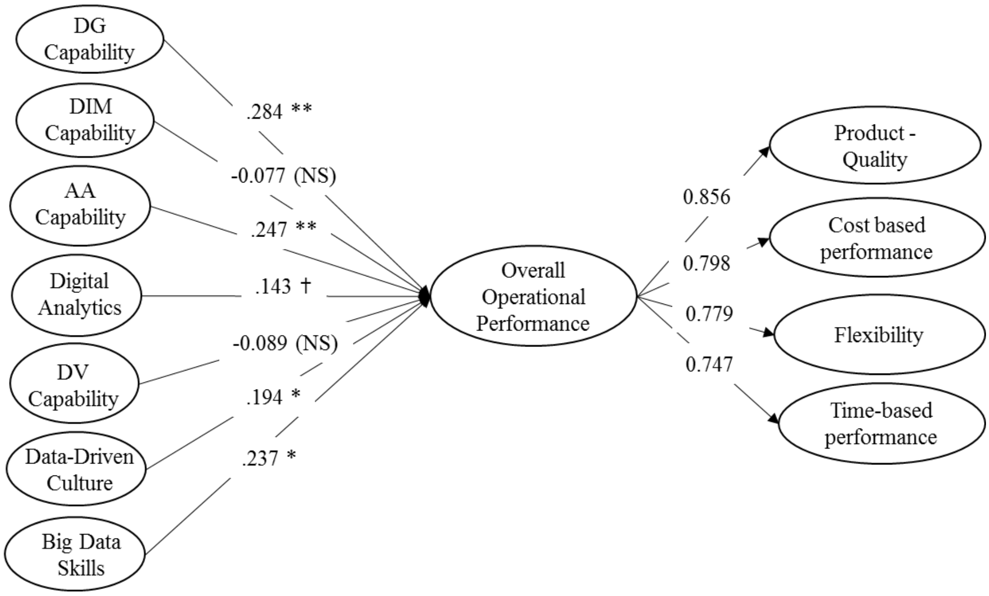


Figure 6.3 Relationship between first-order BDA capabilities and operational performance

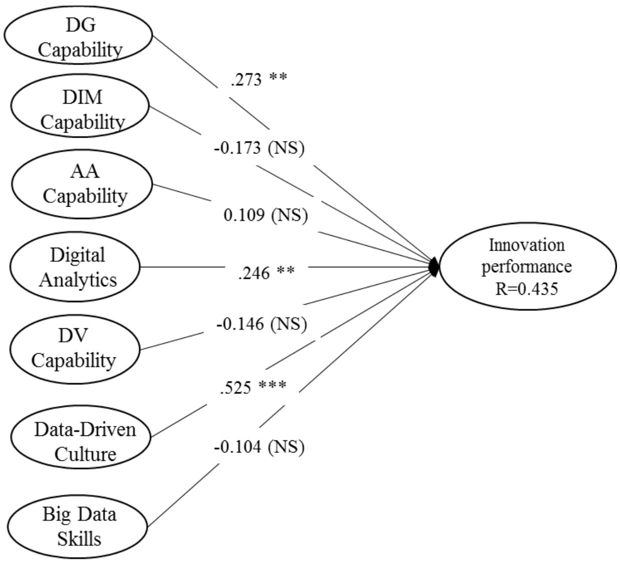


Figure 6.4 Relationship between first-order BDA capabilities and innovation performance

On the other hand, unlike hypothesised, data integration and management and data visualisation capabilities are not found to positively influence operational performance (Figure 6.3). Similarly, data integration and management, data visualisation and big data skills are found to have no significant positive effect on improving innovation performance of manufacturing firms. Previous literature has illustrated the positive influence of data integration on operational performance (Popovič et al. 2009). However, this research does not find enough evidence to support the claim. The 'DIM' capability is predominantly a back-end supporting function for BDA operation and may perhaps be perceived as less significant than the other capabilities such as advanced analytics. Further, it can be argued that data integration and management activities are time-consuming, and the investment and practice of advanced database systems such as NoSQL, and Hadoop are not cost effective due to the requirement for skilled data engineers, which might be reflected in the study's findings. However, considering the effect size and the level of non-significance, DIM capabilities can still be considered as an important capability for enhancing performance. Similarly, the effect of data visualisation capability on operational and innovation performance is also non-significant. Although the capability to visualise data-driven insights intuitively plays a significant role in enhancing the experience of data users, it is found to not directly influence operational performance. A plausible explanation could be that data visualisation may directly influence the decision-making effectiveness (Cao et al. 2015), which, in turn, may influence operational performance. Moreover, it can be argued that a data visualisation capability represents just the delivery of insights, but the real value is extracted from the data by utilising key capabilities such as data generation, advanced analytics, and data-driven culture. Nevertheless, data visualisation is still an important capability to possess to enhance user experience and effectiveness of decision-making. Further, with regard to the non-significant influence of big data skills on innovation performance, it is well-known that innovation is context-specific and requires absorption and deployment of tacit and explicit knowledge for the recognition of business value (Wang et al. 2011). Big data skills represent the employees' ability to perform BDA tasks and these employees might not be directly involved in the innovation process and so could not play a significant role in enhancing innovation. However, as a supportive capability, big data skills are very essential to seamlessly execute data and analytics activities and support decision-makers of the manufacturing organisations.

Moreover, similar to the suggestions of strategic management researchers (Barney and Hesterly 2012b; Mata et al. 1995), this study finds that intangible resources such as data-driven culture are the main source of competitive advantage as it is unlikely to be attained and imitated by competitors. Since firms are made of a bundle of tangible and intangible resources, it is improbable for a firm to compete relying on a single intangible resource. Firms must find ways to integrate and configure tangible and intangible resources to form integrated capabilities such as BDA capabilities maturity. Accordingly, this study has confirmed that the integrated BDA capabilities maturity, composed of tangible and intangible assets of manufacturing organisations, can very likely create sustainable competitive advantage and lead to performance variations between firms in the same industry.

Thus, this study's findings indicate the necessity to develop the aforementioned BDA capabilities to enhance operational and innovation performance of the manufacturing firms. These findings of this study depicting the relationship between BDA capabilities and operational and innovation performance, are in line with some of the previous studies (Prescott 2016; Lee et al. 2014; Lee 2018; Manyika et al. 2011; Gunasekaran et al. 2017; Wamba et al. 2017; Gupta and George 2016; Popovič et al. 2009; Olszak 2016; Lukman et al. 2011; Cao et al. 2015; Gudfinnsson et al. 2015; Jamehshooran et al. 2015b; Warth et al. 2011). The possibility of making a good decision is directly proportional to the veracity of the manufacturing data (Bauer et al. 1994). Ji and Wang (2017) illustrated, using a case study, the benefits of utilising BDA for managing manufacturing shop floor activities. The authors found that BDA can help in identifying high-risk tasks and predict machine faults, which, in turn, can be beneficial for on-time delivery, better utilisation of machines and reduce the cost of extra machining. Soban et al. (2016) demonstrated the use of visual analytics and optimisation algorithms for understanding complex behaviour and optimisation of manufacturing processes. Similarly, by analysing operational data using association rule mining techniques, manufacturing organisations can optimise production schedule for effective utilisation of resources and reduce cost (Li et al. 2015).

Increase in the manufacturing costs is due to the increase in the price of raw material, labour and energy supply. Organisations are highly exposed to price competition in the global marketplace. As a remedy, manufacturers, for instance, automotive original equipment manufacturers (OEM) adopted cost reduction strategies such as lean manufacturing, Toyota Production Systems and instant inventory management systems (used as a measure to reduce supply chain cost). However, these traditional strategies are associated with risks such as minimisation of robustness and supply chain shut down when faced with environmental turbulence. In automotive manufacturing companies, around 60% of total vehicle cost comes from the aggregation of material cost, development and validation cost. Since the traditional processes such as lean manufacturing and six sigma are reactive in nature, they only able to marginally reduce the total cost (Ge and Jackson 2014). With respect to the impact of BDA on cost performance, the findings of this research is similar to the arguments of Ge and Jackson (2014), who argued that Big Data technologies would help the organisation to reduce cost in the automotive industry. The use of Big Data technologies in the remanufacturing process would significantly reduce development and validation cost. Visualisation of batch processing data in manufacturing has a significant impact on monitoring and detecting faults (Wang et al. 2017). Shin et al. (2014) discussed the potential use of BDA in predicting energy consumption and reduce energy cost in manufacturing cost. Similarly, Chien et al. (2016) illustrated the usage of an analytics model to conserve energy and create sustainable manufacturing operations. Application of statistical machine learning techniques such as Support Vector Machines (SVM) and Artificial Neural Networks (ANN) to analyse machine data can improve product yield in semiconductor manufacturing (Lee et al. 2015).

The importance of BDA capabilities for improving manufacturing performance can be highlighted by reiterating a scenario discussed in (Seog-Chan and Hildreth 2016, p.118). A plant manager considers utilisation of a large amount of energy as a mandatory expense for operating manufacturing facilities. However, plant managers intend to receive discounts on the price of energy. In order to do so, an accurate estimate of energy usage over a period of time is required. An overestimation of energy usage will lead to an excess of ‘unused energy’ that is non-refundable. On the other hand, if manufacturers underestimate the energy usage it will lead to the purchase of energy at a non-discounted rate. In either case, the time and efforts put in place to estimate energy could cause financial burdens and budget reconciliation. Traditionally, to determine the energy requirements, multiple regression analysis is used. This technique is only reliable in a manufacturing plant that is stable and functions with minimal variations in processes. In the case of a stable manufacturing environment, the traditional techniques could be beneficial for forecasting energy usage and determining the budget requirements. However, in an uncertain environment, the production facility goes through major changes and significant variance in production processes can be seen throughout the year. These changes may include the addition of new processing equipment, new raw material processing, etc. Traditional regression-based methods are limited in accurately forecasting energy consumption, and to address these issues an advanced activity-based energy accounting (ABEA) technique is developed. This method relies on the acquisition of real-time energy usage data from sub-meters connected to various activities associated with the production process. Moreover, this technique utilises information on the levels of energy required for different levels of production activities such as full-production, reduced-production, and non-production. To continuously collect real-time data, process and analyse it seamlessly, it is essential for the manufacturing firms to possess BDA capabilities. Thus, it is evident from the findings of this research and the above-mentioned case scenario, by combining the real-time energy usage data with advanced predictive analytics techniques, manufacturing plants could accurately predict the energy consumption, thereby reducing cost and enhance sustainability practices. Overall, the findings of this study revealed sufficient evidence that BDA capabilities can significantly improve operational and innovation performance.

6.3.2 Relationship between BDA maturity and absorptive capacity, data and information quality, and supply chain analytics capability.

From the perspective of the hierarchy of capabilities, scholars have suggested that lower-order capabilities can facilitate firms to build higher-order capabilities (Kraaijenbrink et al. 2010; Grewal and Slotegraaf 2007; Grant 1996a; Sirmon et al. 2007). In this research, it is hypothesised that BDA capabilities maturity, a lower-order capability, influences firms to develop higher-order capabilities such as absorptive capacity (ACAP), data and information quality (DIQ), and supply chain analytics (SCA) capability. Consequently, based on the SEM, it is found out that BDA maturity has a significant positive effect on ACAP (0.684***), DIQ (0.688***), and SCA (0.673***) capability (Figure 6.5).

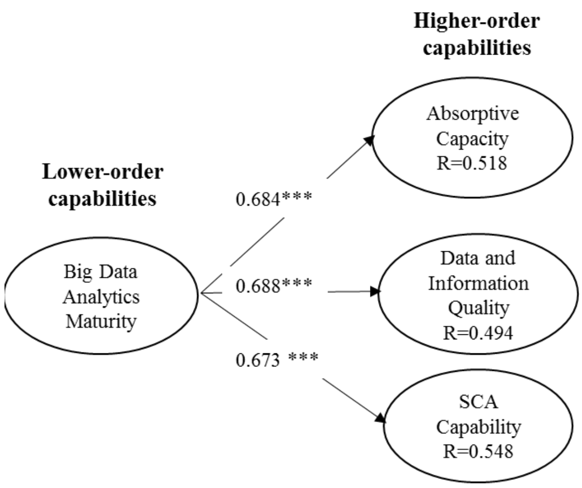


Figure 6.5 Relationship between lower-order capabilities and higher-order capabilities

In line with previous studies, the current study reinforces that the lower-order capabilities enhance firms' ability to develop higher-order capabilities (Sambamurthy et al. 2003; Liu et al. 2013). First, with regard to the impact of BDA on ACAP, the current study has confirmed that BDA capabilities are an antecedent of absorptive capability, which is in line with studies such as (Malhotra et al. 2005; Wang et al. 2014; Fayard et al. 2012; Noblet et al. 2011; Ettlie and Pavlou 2006). Accordingly, BDA can improve the ability of organisations to manage knowledge acquired from the external environment and effectively use it to generate a commercial advantage. Improvement of BDA capabilities increases the flow of information and knowledge which can have a positive effect on the R&D initiatives of a firm, that are crucial for innovation (Bellamy et al. 2014). Further, it is confirmed that firms with enhanced lower-order organisational capabilities such as BDA maturity can significantly improve higher-order dynamic capabilities such as ACAP. A firm may be able to recognise valuable knowledge in the external environment but could perhaps face significant difficulties in acquiring and assimilating it within the organisation. When firms struggle to incorporate new knowledge, the richness of knowledge structures gets affected (Cohen and Levinthal 1990). However, as suggested by this research, firms can combine various BDA-related capabilities to enhance the acquisition, manipulation, and interpretation of knowledge received from external partners, thus enhancing the assimilation of knowledge throughout the organisation. BDA capabilities could allow firms to collect a vast amount of knowledge from their external partners, but firms should develop mechanisms to recognise valuable information and facilitate its flow within the organisation.

Second, regarding the impact of BDA on data and information quality, this study has confirmed that higher levels of BDA maturity will improve DIQ of organisations. This finding is in line with the previous studies (Strong 1997; Popovič et al. 2012; Popovič et al. 2009; Nelson et al. 2005). In particular, Hannula and Pirttimäki (2003) and Popovič et al (2012) have identified that an increase in the quality and maturity of Business Intelligence System will influence the data and information quality. For decision making, organisations prefer to use data and information that are of high quality, and from easily accessible sources (Popovič et al. 2012; Hannula and Pirttimäki 2003). With BDA, DIQ can be enhanced by clearly defining the data structures within the data management systems. Master data management functionality and data governance mechanisms embedded within the big data warehouse systems will play a major role in ensuring data and information quality. The practice of BDA ensures that the data is clean, secure and free from errors, and makes it reliable for decision-making. Therefore, it is imperative for organisations to pay close attention to utilise BDA to perform data integration and management activities intended to ensure data and information quality.

Third, similar to results of the impact of BDA capabilities maturity on ACAP and DIQ, this research has found significant evidence to confirm that the maturity of BDA capabilities (at intra-organisational level) can improve supply chain analytics or data-driven supply chain capabilities. BDA maturity explains more than 50% of the variance in the supply chain analytics construct, which verifies that intra-organisational BDA maturity is a strong predictor of SCA capabilities. This finding is in line with previous research which claims that organisations' experience and comfort of deploying a technology or practices at intra-organisational level, can significantly enable them to deploy their capabilities at inter-organisational level (DeGroote and Marx 2013; Hefu Liu et al. 2013b). The know-how of BDA practices and the awareness of the potential benefits of using it for managing supply chain activities such as sourcing, supply network design, and logistics optimisation could strongly encourage firms to utilise it beyond the organisational level. Similar to the views of Luzzini et al (2015), this research finds that, in order to gain a greater competitive advantage, organisations are committed to developing both intra- and inter-organisational capabilities. Further, in line with the findings of (Hefu Liu et al. 2013b), this study confirms that an increase in the level of lower-order organisational capabilities, can influence firms to develop higher-order operational capabilities such as SCA capability. Thus, it can be argued that an organisation's ability to reconfigure lower-order capabilities has a greater influence too on various functional areas such as supply chain. This kind of resource reconfiguration and utilisation denotes a prior knowledge base that is tacit in nature and difficult to imitate by competitors. Hence, this study confirms that an organisation's increased level of expertise in BDA practice for intra-organisational decision-making is the antecedent of data-enabled supply chain practices.

6.3.3 The mediating role of absorptive capacity on the relationship between BDA and operational and innovation performance

This section discusses the results for the hypotheses relating to the mediating effect of ACAP on the relationship between operational and innovation performance. As noted in (Lane et al. 2006), Absorptive Capacity (ACAP) should be empirically researched in various business contexts other than the Research and Development(R&D), where it is substantially researched. This research explores the importance of ACAP in the context of BDA capabilities and business value creation. As suggested by Setia and Patel (2013) and Liu et al. (2013), the applicability of ACAP as a consequent of technological capabilities is investigated, the results of this study found evidence on the significant role of ACAP in the context of this research. Based on the analysis of the mediation model, it is found that a higher level of absorptive capacity partially mediates the positive relationship between: 1. BDA maturity and product quality (0.595*** without, and 0.289*** with ACAP as a mediator), 2. BDA maturity and Cost-based performance (.615*** without, and 0.336*** with ACAP as a mediator), 3. BDA maturity and flexibility (.554*** without, and 0.347*** with ACAP as a mediator), and 4. BDA maturity and Time (.610*** without, and 0.378*** with ACAP as a mediator), as shown in Figure 6.6. Moreover, considering the mediating effect of ACAP on the relationship between BDA maturity and overall operational performance, results indicated the presence of partial mediation (0.599*** without, and 0.281*** with ACAP as a mediator). Similarly, as hypothesised, this study has found evidence to support the mediation effect of ACAP on the relationship between BDA maturity and innovation performance (0.569*** without, and 0.419*** with ACAP as a mediator). The indirect effect of BDA maturity on product quality (0.269***), cost (0.241**), flexibility (0.187*), time (0.221*), and innovation performance (0.304***) are found to be significant.

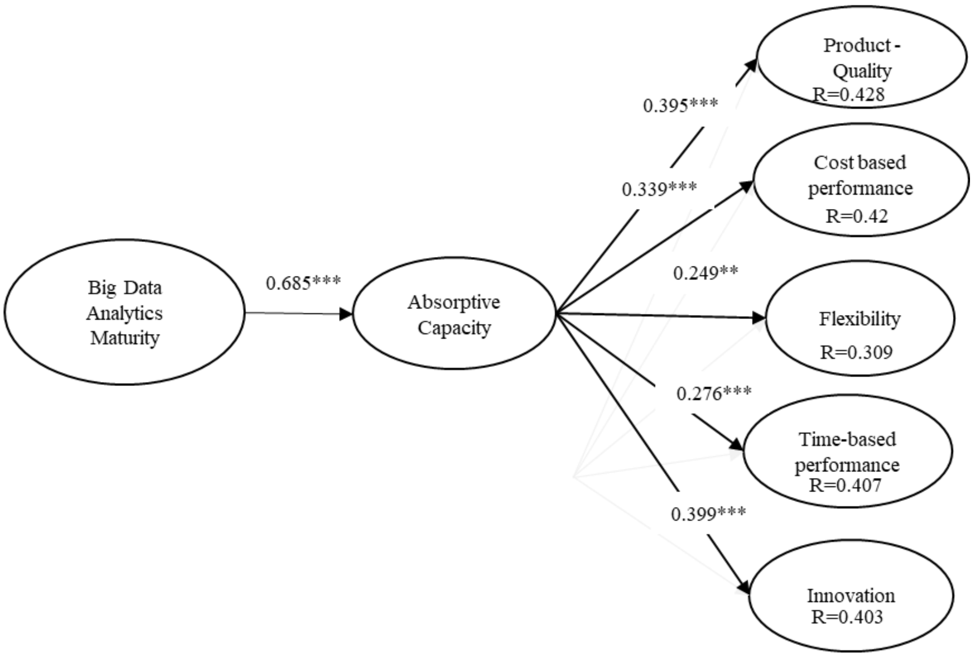


Figure 6.6 Relationships of BDA maturity, ACAP, operational and innovation performance

In line with previous studies (Ali and Park 2016; Jabar et al. 2011), this study confirms that ACAP can significantly influence an organisation's product, process, and management innovation. ACAP, as a dynamic capability, has the potential to enhance organisational learning mechanisms acquiring, assimilating, transforming, and applying external knowledge, and this, in turn, supports new product development in a manufacturing context. ACAP is also regarded as a prior knowledge possessed by workers and plant managers and, their ability to share the knowledge is crucial for the implementation of innovative management practices (Tu et al. 2006). Similar to the views of Tu et al. (2006), this study finds that ACAP significantly improves the firm's ability to innovate new management practices.

Additionally, analogous to the findings of Beheregarai Finger et al. (2014), ACAP is found to have a significant role in enhancing quality, cost, flexibility and time dimensions of operational performance. In particular, Fayard et al. (2012) claimed that ACAP improves the cost-based performance of manufacturing by indirectly influencing inter-organisational cost-management practices. Similar to Fayard et al. (2012)'s observation, this research has confirmed the role of ACAP in improving cost-related performance. Consistent with Patel et al. (2012) and Tu et al. (2006), this study finds that ACAP influences a firm's ability to carry out time-based manufacturing practices to improve performance. In addition, firms with better ACAP are more likely to respond proactively to environmental uncertainties and improve manufacturing flexibility.

In the broad sense, if a manufacturing organisation is willing to improve operational and innovation performance then a certain level of information exchange with the external environment is needed. However, the extent of utilising external knowledge for the benefit of the business depends on the ability of organisations to acquire, assimilate and apply the valuable external knowledge. BDA capabilities can be viewed as a supportive mechanism crucial for knowledge exchange. Some of the support activities that can be carried out by utilising BDA include acquiring information from the external source and processing it and making it uncomplicated for the employees to exploit such information. The knowledge exchange mechanism created by BDA capabilities enhances firms' ability to assimilate the external knowledge with prior internal knowledge-base, transform and apply it for commercial purposes such as for improving the efficiency of product/services offerings, innovating new products/services to increase customer satisfaction. Thus, as highlighted by Cohen and Levinthal (1990), organisations must develop ACAP in order to fully utilise the flow of external knowledge facilitated by the adoption of BDA capabilities. Overall, the examination of the mediating role of ACAP has provided sufficient evidence that the development of BDA capabilities and utilising it to improve operational and innovation performance, requires firms to focus more on improving their ACAP.

6.3.4 The mediating role of data and information quality on the relationship between BDA and operational and innovation performance

This section discusses the test results of the hypotheses relating to the mediating effect of data and information quality (DIQ) on the relationship between operational and innovation performance. Based on the path analysis of the mediation model, it is found out that a higher level of Data And Information Quality (DIQ) partially mediate the positive relationship between: 1. BDA maturity and product quality (0.595*** without, and 0.439*** with DIQ as a mediator), 2. BDA maturity and Cost-based performance (.615 *** without, and 0.408*** with DIQ as a mediator), 3. BDA maturity and flexibility (.554 *** without, and 0.419*** with DIQ as a mediator), and 4. BDA maturity and Time (.610 *** without, and 0.40*** with DIQ as a mediator). Moreover, considering the mediating effect of DIQ on the relationship between BDA maturity and overall operational performance, the results indicated the presence of partial mediation (0.569*** without, and 0.45*** with DIQ as a mediator). However, in contrast to the proposed hypothesis, this study did not find evidence to support the mediation effect of DIQ on the relationship between BDA maturity and innovation performance as the indirect effect is non-significant (0.138(NS)).

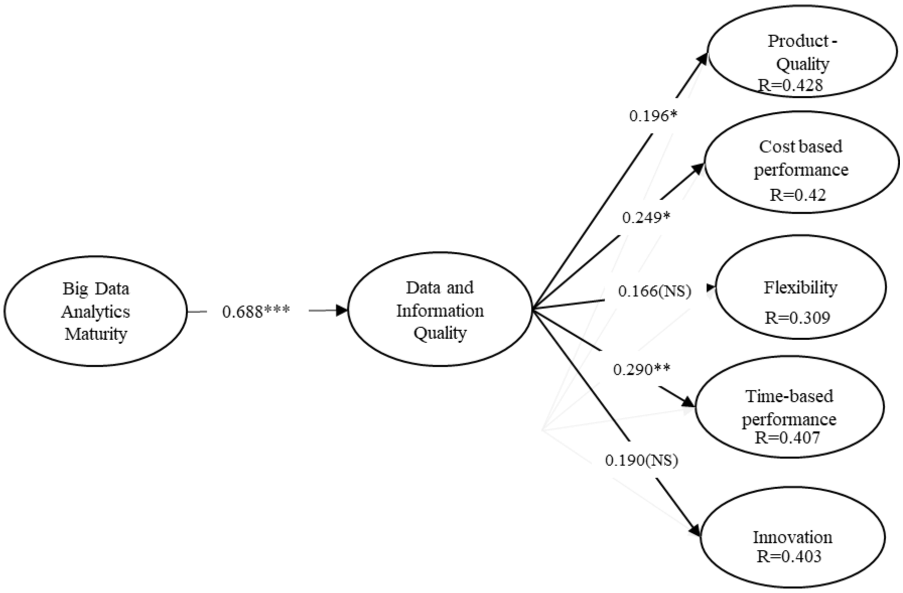


Figure 6.7 Relationships of BDA maturity, DIQ, operational and innovation performance

In line with previous studies (Woodall et al. 2013; Redman 1998; Redman 2001; Redman 1996; Haryadi et al. 2016; Pipino et al. 2002; Wang and Strong 1996; Shen et al. 2015; Kwon et al. 2014; Song et al. 2017; Abdullah et al. 2015; Hazen et al. 2014; Chae, Yang, Olson, et al. 2014), data and information quality is found to influence product quality, cost and time-related operational performance, except flexibility. However, in contrast to the proposed hypothesis but consistent with the views of Kim et al. (2012), this study has found that the quality data and information has no significant effect on innovation. Perhaps it can be argued that the availability of DIQ alone cannot directly influence some of the operational activities and innovation. As discussed in chapter 2 (section 2.8), innovation is a complex process that involves dynamic utilisation of resources to enhance innovation. DIQ is a higher-order organisational capability (static in nature), which may have a direct influence on the effectiveness of decision-making (Malhotra et al, 2005; Wiengarten et al. 2010), but its effect is not reflected on the innovation performance of the manufacturing organisation. However, similar to the findings of Kaynak (2003), this study has found evidence to justify the claim that the availability of DIQ can enhance the manufacturing process such as monitoring of product quality and accuracy of product delivery. Maintaining DIQ can help workers in the manufacturing plant to accurately monitor any deviation in the processes and quality requirements in real-time during the production process. Moreover, in line with Setia et al. (2013) and Nelson et al. (2005), this study has found significant evidence of the positive influence of DIQ on on-time delivery related performances, this, in turn, can increase customer satisfaction. Although this study did not find evidence to support the positive influence of DIQ on manufacturing flexibility, based on Gustavsson and Wänström (2009) it is evident that the availability of timely and reliable information will decrease the production lead time and improve flexibility. Also, the completeness, reliability and accessibility aspect of DIQ can improve responsiveness and better forecasting. With regard to the role of DIQ on cost performance, in line with Selvage et al (2017), Gustavsson and Wänström (2009), Lee et al. (2006), Redman (1998) and Redman (1996), this study found out that the ability to maintain DIQ can reduce inventory cost, manufacturing cost and overhead cost and enhance the competitiveness of manufacturing firms based on price. While poor DIQ can cost organisations an average of 8-12 per cent of the revenue, efforts to maintain DIQ can certainly reduce these cost and create a competitive advantage as a result of improved cost performance. Moreover, as contented in Chae, Yang, Olson, et al (2014), Kim et al. (2012) and Krause et al. (1998), this study has found evidence that the availability of high-quality data and information about supplier performance integrated from a supplier's database, can support manufacturing firms to proactively track and solve problems that may arise from supplier side and as such increase accuracy of planning.

Similar to the views of O'Donnell et al (2012), this research also shows that data management should be one of the top priorities among BDA practitioners, and much of their efforts should be focused on data quality. Consistent with Watson and Wixom (2007), this research highlights the importance of maintaining data quality as a key to the success of BDA implementation and practice. For instance, analogous to the illustration of Gudfinnsson et al. (2015), a manufacturing organisation having sales offices distributed in various locations will be benefited by the process of acquiring and integrating data from the distributed databases and achieve the status of 'data completeness'. While integrating data may ensure the trustworthiness and reliability of data, it is also important to make sure that the data available for all employees of the organisation provide a consistent representation of the reality, signifying ‘data consistency’. Moreover, it is necessary to ensure that data and information are not delayed and easily accessible by the data users, signifying ‘data accuracy/accessibility’. Thus, like the expression “garbage in, garbage out”, poor data and information quality always will end up in poor decisions. Therefore, it can be noted from this study that improved DIQ can help manufacturing firms achieve sustainable competitive advantage and improve performance through decreasing manufacturing costs, improving customer satisfaction, and eliminating non-value-added activities, but it does not have any direct influence on innovation.

6.3.5 The mediating role of SCA capability on the relationship between BDA and operational and innovation performance

Management of information flow is one of the core supply chain activities and its importance is increasing since firms in the supply chain strive to be more responsive to intensifying customer demand for products that are of high quality and innovative in nature. However, effective processing of supply chain information remains a challenge. The development of data-driven supply chains could overcome the challenges by creating a more proactive supply chain.

This section discusses the results for the hypotheses relating to the role of SCA capability (or data-driven supply chain capability) on the relationship between operational and innovation performance. Based on the path analysis of the mediation model, it is found that a higher level of SCA capability partially mediates the positive relationship between: 1. BDA maturity and product quality (0.595*** without, and 0.208* with SCA as a mediator), 2. BDA maturity and Cost-based performance (.615*** without, and 0.263* with SCA as a mediator), 3. BDA maturity and flexibility (.554*** without, and 0.272* with SCA as a mediator), and 4. BDA maturity and Time (.610*** without, and 0.276**with SCA as a mediator). Moreover, considering the mediating effect of SCA on the relationship between BDA maturity and overall operational performance, results indicated the presence of partial mediation as well (0.569*** without, and 0.276**with SCA as a mediator). However, in contrast to the proposed hypothesis, this study did not find evidence to support the mediation effect of SCA on the relationship between BDA maturity and innovation performance because the indirect effect is non-significant (0.198(NS)).

These findings are in line with existing studies (Yu et al. 2017; Gunasekaran et al. 2017; Sangari and Razmi 2015; Sanders 2016; Jamehshooran et al. 2015a; Hazen et al. 2014; Davenport 2006; Kache and Seuring 2017; Chircu et al. 2014; Chae, Yang and Olson 2014), and this study suggests that effective management of supply chain activities via data-driven practices could improve manufacturing processes and performance. The results also indicate an important link between BDA maturity and SCA capabilities (β=-0.673***) and further, it is much stronger than the relationship between BDA maturity and product quality (β=-0.208*), cost (β=-0.263*), flexibility (β=-0.272*), time (β=-0.276**) and innovations (β=-0.338**). Alternatively, the positive relationship between BDA maturity and SCA capability further enforces substantial influence on operational performance. It can be argued that SCA plays an important role in achieving operational excellence through the antecedent of BDA maturity, although BDA maturity does indeed have a positive impact on operational performance. Comparatively, the structural path indicating the mediation of SCA capability is the significant contributor to operational performance. One possible explanation for this scenario is that BDA practices may be treated as a strategic action and firms try to prepare themselves for nurturing their BDA capabilities for establishing data-driven SCM practices. BDA, indeed, plays the role of a key enabler of data-driven supply chains. This may involve building a customised BDA infrastructure for the purpose of managing supply chain activities. Next, while BDA maturity is considered as a lower-order organisational capability enabling SCA (a higher-order operational capability), they are both important for influencing operational performance but at varying magnitudes. As the practice of BDA at intra-firm level becomes more sophisticated, organisations will tend to utilise it to manage inter-firm activities such as strategic sourcing and supply network optimisation. As manufacturing firms get more proficient with BDA practices for internal and external purposes, they will produce products that comply with quality requirements, reduce costs and increase delivery and time-related performances, and this will further facilitate them to focus more on their core competencies as both BDA maturity and SCA are antecedents of operational performance.

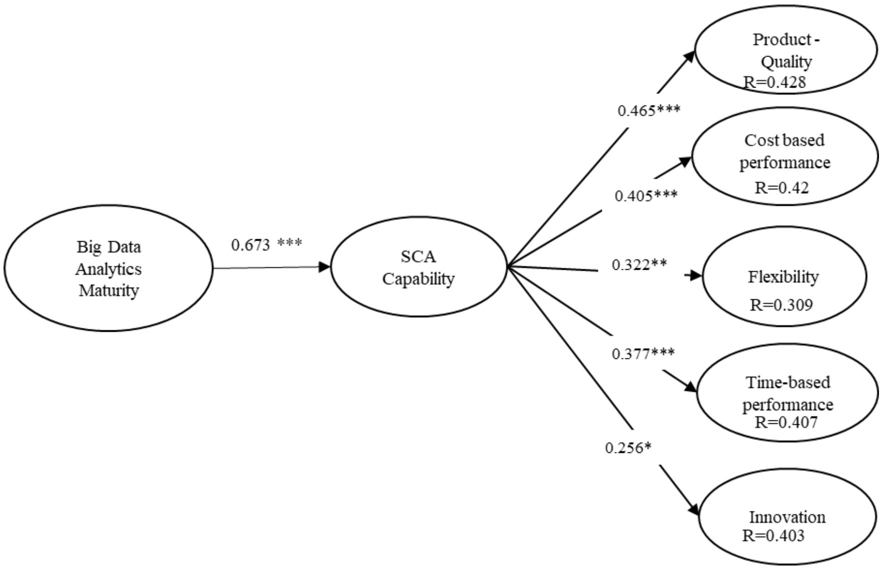


Figure 6.8 Relationships of BDA maturity, SCA capability, operational and innovation performance

Specifically, concerning the influence on quality performance, in line with Davenport (2006), this study finds that the exchange of data from suppliers and customers in real-time could reduce product defects via prototype testing and simulation of products at the early stage of development. Consistent with Dubey et al. (2015), Lee et al. (2013) and Wamba et al. (2017), this study shows that the characteristics of data-driven supply chains would allow firms to reduce inventory level as well as reduce the probability of being out of stock at the supply chain level, thus reducing cost. Similarly, with regard to the time-related performances, our findings are consistent with the views of Manyika et al. (2011), Brynjolfsson and McAfeeErik (2012) and Wamba et al. (2015), that data-driven supply chains enable collaborative practice across the supply chain and can reduce product development time, on-time delivery, and delivery speed by enhancing visibility and optimisation of supply chain flow. Also, the findings of this research corroborate the views of Sanders (2016) and McAfee and Brynjolfsson (2012), by being aware of the disproportionateness of demand and the supply capacity, and recognising the needs and demands of customers organisations can be more flexible and quickly respond to unexpected changes in demand and supply. Further, SCA enables the availability of even the granular level information about supply chain operations which allows manufacturing firms to make more customisations in terms of volume.

Moreover, in the presence of SCA, the indirect effect of BDA maturity is significant only on operational performance, but it does not have a significant effect on the innovation performance. Theoretically, it is anticipated that the practice of data-driven SCM could enhance information exchange from competitors, customers, and suppliers, which in turn, might positively influence innovation performance of manufacturing. However, in contrast to the hypothesis, this research finds SCA does not play a significant role in the relationship between BDA maturity and innovation performance. This may be because most large firms are more proficient in utilising SCA capabilities, and when the focal firms cannot build effective mechanisms to support data-driven supply chain activities with their partners, it may not be possible to collaborate with partners in activities such as the design of innovative processes and the development of new products. Also, information exchange from external and internal sources occurs in an open system which may enable firms to innovate products and processes. Perhaps, in the context of the UK manufacturing firms, open innovation may be more common in high-value manufacturing firms than in traditional firms. Still, there is a significant direct effect of SCA on innovation. Another plausible explanation of this scenario is since ACAP plays as a strong mediator of BDA maturity and innovation performance, SCA as an operational capability has less influence and also it is predominantly used “to make a living in the present” (Brusset 2016, p.47). Hence, utilising BDA tools and technologies as a complementary support system to enable traditional SCM practices can improve operational performance. However, a flexible BDA infrastructure is necessary to establish a physical link between members of the supply chain so as to make information exchange more feasible for managing product and finance flow activities. Organisations should be aware of the fact that the social and technological concerns of supply chain partners have to be recognised to realise the full potential of creating a data-driven supply chain to enhance operational as well as innovation performance.

6. 4 Exploring the phenomenon of the digital divide between SMEs and Large organisation

In this section, the findings from the clustering analysis used to observe the phenomenon of the digital divide between SMEs and large organisations are discussed. By utilising various cluster analysis techniques, the four cluster structure hidden in the dataset is revealed and the characteristics of these four clusters are identified using cross-profiling technique and discussed in the following sections.

6.4.1 Cluster 1 – Routine users / Advanced users

The key characteristics of this cluster are: 1. 85% of companies that belong to this cluster have more than 100 employees, and 2. Most of their annual turnover is greater than 25 million pounds (Figure 6.10). Compared to other clusters, this cluster is dominated by large organisations but also includes a few small and medium-sized manufacturing firms. However, this is one of the smallest clusters with only 18% of the companies who participated in this study are being categorised into this cluster.

In terms of BDA capabilities, the organisations in this cluster are comparatively more highly matured than those in the other clusters identified in this research (Figure 6.9). It can be argued that the members of this cluster are the strongest in developing and utilising BDA capabilities. These companies exhibit the ability to generate data using a variety of tools such as sensors and IoT. Moreover, they are leading in terms of utilising data integration tools such as ETL and clearly demonstrate the ability to use advanced database technologies such as Hadoop, MapReduce, and NoSQL. Besides, they are the best in the usage of advanced analytics techniques such as association rule mining, neural networks, and machine learning, as well as leaders in the utilisation of social media and web analytics. The usage of BDA in the supply chain function is also relatively higher than in the other identified clusters. Their ability to use digital analytics and supply chain analytics indicates their focus on collecting, integrating and analysing external data from suppliers and customers. High levels of data visualisation capability indicate that the organisations in this cluster do not rely on static reports for the delivery of information, rather data and information are delivered using interactive tools such as dashboards and web-based application systems. In addition, these organisations seem to be the strongest in recruiting and training employees capable of performing data and analytics tasks such as data scientists, data engineers, and business analysts to support data-driven decision-making. In the structural path analysis, data-driven culture is identified as an important capability to enhance firm performance. Data-driven culture reflects the attitudes and beliefs of top management and employees in the organisation and the member of this cluster greatly consider data as a strategic asset for decision-making and are willing to rely on it along with their expert judgment.

Due to the strong presence of essential BDA capabilities, especially data integration and management capabilities, organisations that belong to this cluster can maintain data and information quality. This enables the data users and decision makers of these organisations to have access to high-quality information content in a timely manner, which ultimately improves the effectiveness of decision-making. These organisations can efficiently utilise data and information to monitor business processes, quickly react to the changing environment, and customise new products and services. Apparently, organisations in this cluster are more advanced in utilising BDA and can be considered as 'data rich' and 'information rich'. From RBV perspective, it can be argued that the members of this cluster possess a unique set of tangible, intangible and human resources that are capable of generating economic rents and ultimately create a sustainable competitive advantage. Thus, based on the characteristics of this cluster, these organisations are named as 'routine users' and distinguished as 'highly mature' organisations.

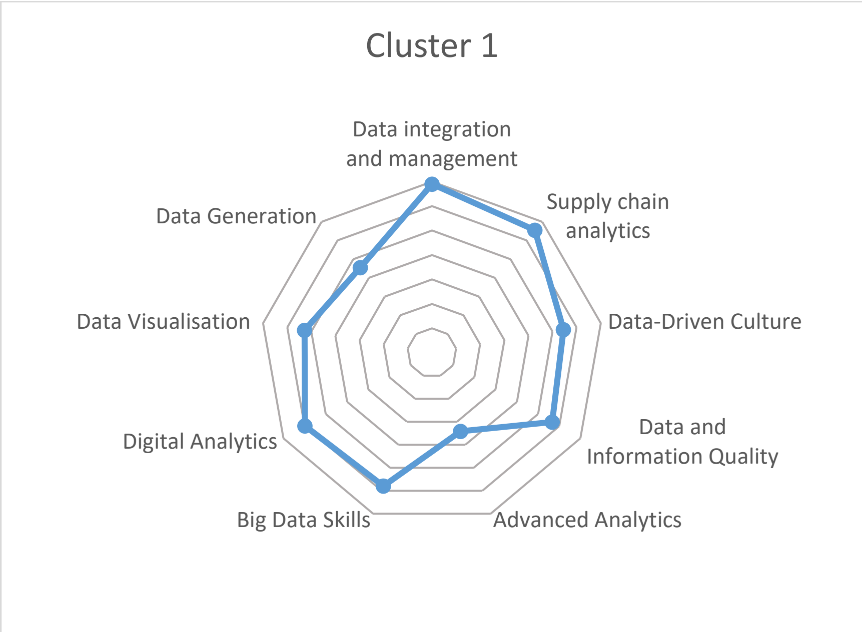


Figure 6.9 Spider plot of cluster 1 Vs BDA capabilities

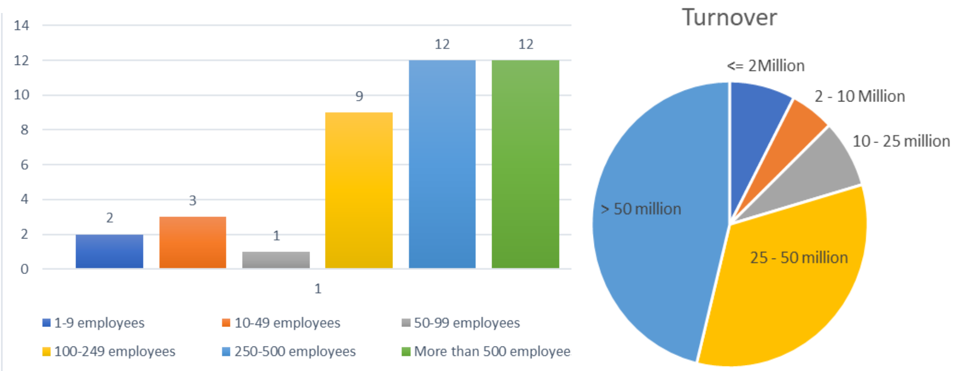


Figure 6.10 a) Cluster 1 Vs Number of employees b) Cluster 1 Vs Annual turnover

6.4.2 Cluster 2 - Ad-hoc users

The key characteristics of this cluster are: 1. 70% of the companies belonging to this cluster have an annual turnover of £10-25 million, 2. with respect to the number of employees, 75 % of them have employees ranging from 49 to 249 (Figure 6.12 b). These characteristics indicate that most of the companies in this cluster are SMEs with a larger proportion of medium-sized firms. Moreover, this is one of the largest clusters identified in this research with 32% of the firms who participated in this study (Figure 6.12 a).

In term of BDA capabilities, the organisations in this cluster have indicated that they possess high levels of advanced analytics capabilities, data-driven culture and data and information quality (Figure 6.11). However, the level of these capabilities is comparatively lower than that of Cluster 1. The members of this cluster can generate data to some extent, but they generally lack data integration and management abilities. It can be argued that these organisations are still managing data using legacy practices such as Relational Database Management Systems (RDBMS) and are not proficient with the utilisation of advanced databases technologies such as NoSQL (Not only SQL) and Spatio-temporal databases to manage data generated in a distributed environment. Since these organisations are not capable of generating high volume and variety of data such as text, social media data and RFID data, the necessity of utilising advanced data integration and management tools is minimal. Further, they lack big data skills, which suggests that their data related tasks are mostly managed by their IT department- if it is existing. However, the stakeholders of these organisations do realise the importance of data-driven decision making, shown by the presence to some extent of a data-driven culture. Also, these organisations seem to do well in terms of maintaining data and information quality. It could be argued that since they generate less complex data, monitoring and maintaining data and information quality is uncomplicated. Moreover, the presence of advanced analytics capabilities and data-driven culture indicate that these organisations are willing to utilise BDA but are in the early stages of adoption. They seem to utilise analytics on an ad-hoc basis to analyse the data but there is no evidence of the institution-wide practice of data-driven decision making. Hence, based on these characteristics, organisations in this cluster are named as ‘ad-hoc’ users. From RBV perspective, it can be argued that they lack the data generation capability (source for creating data), found to be significant for improving firm performance. Relatively, these organisations can be categorised as 'data poor' and 'information poor' as they fail to capture data from various sources, and ultimately would fail to attain competitive advantage. Nevertheless, the characteristics of the organisations in this cluster show that they are slightly more mature than cluster 3 (beginners/laggards), discussed in the next section.

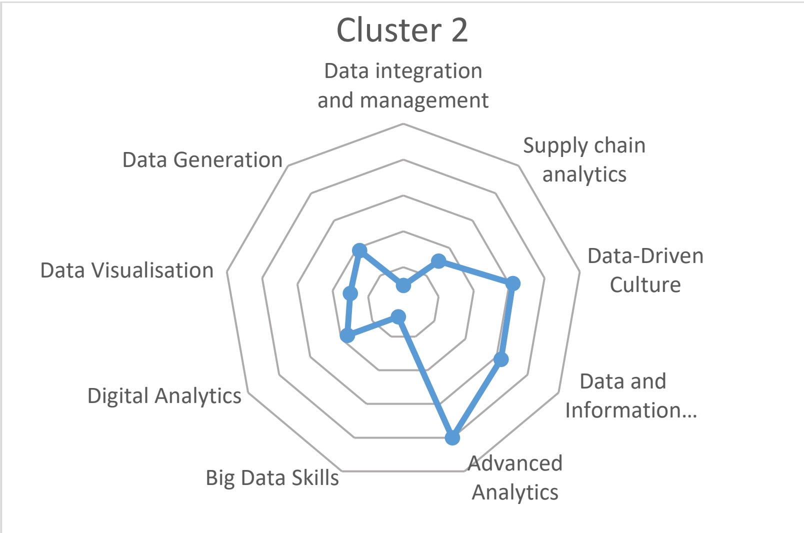
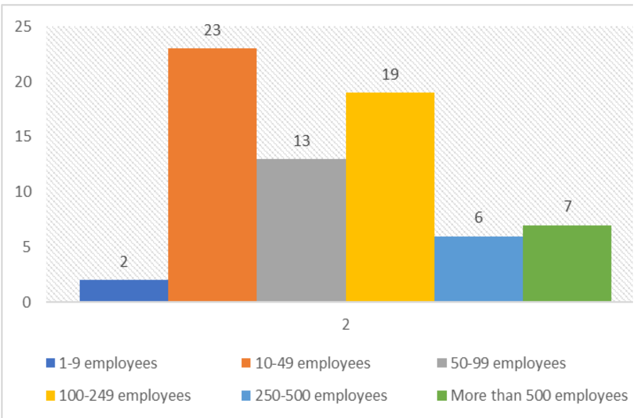


Figure 6.11 Spider plot of cluster 2 Vs BDA capabilities



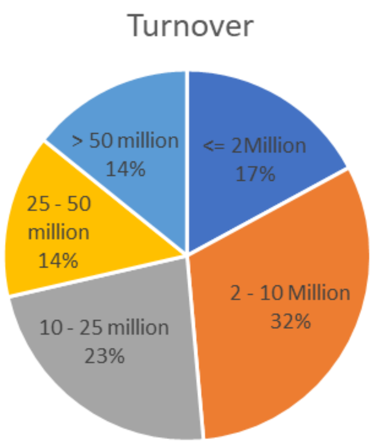


Figure 6.12 a) Cluster 2 Vs Number of employees *b) Cluster 2 Vs Annual turnover*

6.4.3 Cluster 3 –Beginners / laggards

The key characteristics of cluster 3 are: 1. 40% have employees ranging from 10-49 and 23% have 100-249 employees in their organisation, 2. 60% of the companies belonging to this cluster have an annual turnover of £ 2-10 million (Figure 6.14 b). This is probably the third largest cluster with 19% of participating firms belonging to this group. Based on the demographics, the majority of the organisations that belong to this cluster are small in size (Figure 6.14 a).

With respect to BDA capabilities, it is apparent that the main focus of the organisations in this cluster is to generate data and use some form of digital analytics, but their use is relatively lower than the organisations in other clusters (Figure 6.13). Moreover, these organisations use BDA at the supply chain level more than at the intra-organisational level. This is evident from Figure 6.13, their lack of ability to use advanced analytics, data integration and database tools compared with the other clusters identified in this research. As a result, it is reflected in their ability to maintain data and information quality, generally lowest among other clusters as data integration and management capabilities are significant to maintain data quality (Gudfinnsson et al. 2015). Further, to a certain extent, these organisations have an inclination towards the usage of analytics at the supply chain level. This could be an indication of external pressure from their supply chain leaders. For instance, 'Walmart' a supply chain leader insists that their suppliers share data and use data-driven insights for the effective management of supply chain processes (Sanders 2016). Similarly, it can be argued that the UK manufacturing organisation in this cluster could be influenced by their supply chain members to use some form of analytics to support supply chain activities. However, their lack of data-driven culture, top management support, and other critical resources inhibit them from developing key capabilities such as DIM, DVC, BDS and advanced analytics. Based on the characteristics of this cluster, this cluster is named as 'Beginners/laggard'. These organisations are not only at the beginning of the journey of BDA and data-driven decision-making but also relatively 'data poor' and 'information poor'. From RBV perspective, it can be argued that these organisations do not possess unique BDA resources and relatively lack competitive advantage, and hence categorised as 'competitively disadvantaged' organisations.

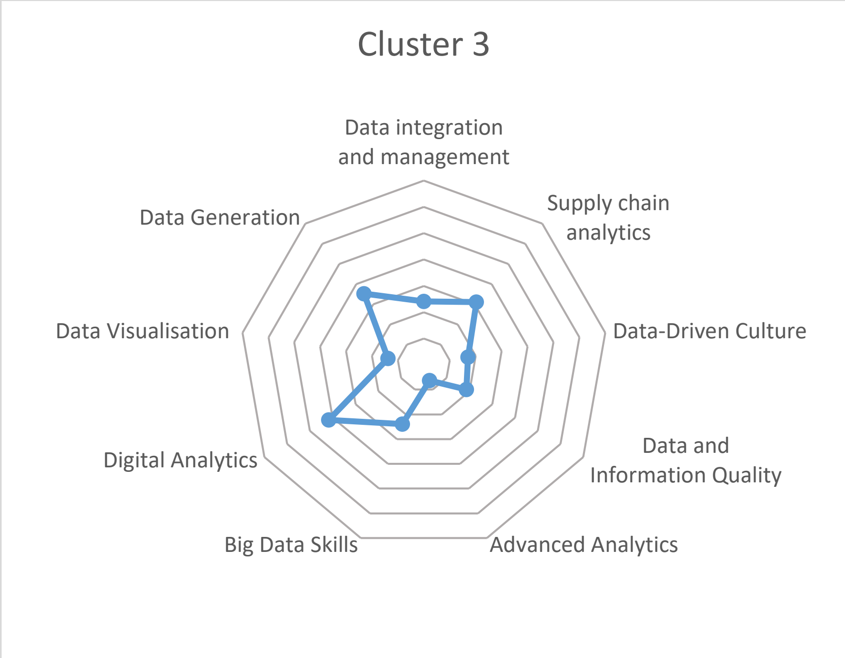


Figure 6.13 Spider plot of cluster 3 Vs BDA capabilities

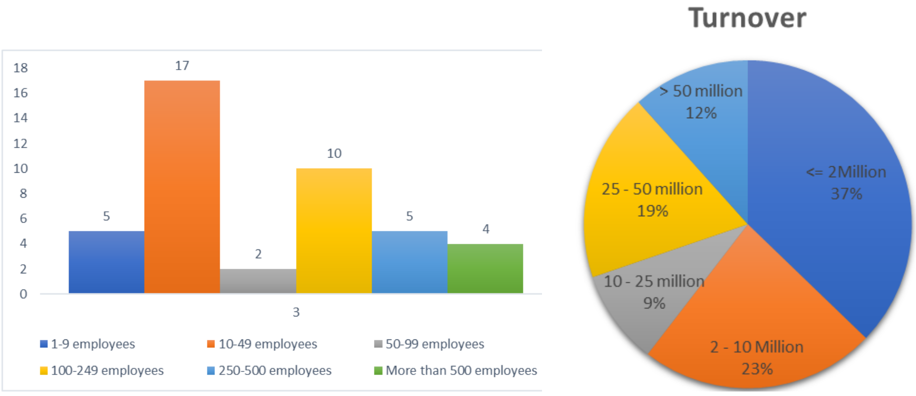


Figure 6.14 a) Cluster 3 Vs Number of employees b) Cluster 3 Vs Annual turnover

6.4.4 Cluster 4 – Enthusiastic users

This is the second largest cluster with 31% of firms who participated in this study belonging to this group. Unlike the other clusters, cluster 4 has a genuine mix of firms with a varying number of employees. For instance, 20% of them have 10-49 employees, 16% of the firms have 50-99 employees, 31% of firms have 100-249 employees, and 24% of firms with more than 500 employees (Figure 6.14 a). However, in terms of turnover, close to 60% of the firms make an annual turnover of £25 million to £50 million and above (Figure 6.16 b). These characteristics of the cluster indicate that the firms that belong to this cluster are mostly medium and large firms but also include a considerable number of small firms.

The members of this cluster group are using a considerable amount of BDA technologies. From the spider plot (Figure 6.15), it is evident that this group is generating a high volume of big data and is relatively ‘data rich’. However, these organisations are different from ‘Cluster 1-routine users’ in several ways. For instance, their data and integration capabilities are comparatively lower than high cluster 1 but bhigher than cluster 2 and 3. This indicates that these organisations are aware of the importance of integrating data from multiple sources using technologies such as OLAP and Extract Transfer and Load (ETL). However, due to some unknown reason, they are not completely integrating data from both internal and external sources. Deficiency in data integration capabilities causes poor maintenance of data quality. Although these organisations show interest in generating an enormous volume of data, their lack of DIM capabilities would inhibit them from extracting full value from the technology due to data quality issues. Further, these organisations are sufficiently equipped with advanced analytics, data visualisation and big data skills, which indicate that they are enthusiastic about developing BDA capabilities for data-driven decision making. However, compared to cluster 1, these organisations do lack a data-driven culture and supply chain analytics capabilities. This indicates that the perception and views of top management are not supportive enough to use the innovative forms of analytics such as digital analytics. Also, these organisations can be considered as more internally-focused as they generate and integrate mostly internal data and use analytics for managing the internal business process, but not willing to use it at the supply chain level. A plausible explanation could be that strategically these organisations perceive data as a valuable source of competitive advantage, but due to privacy and security aspects of big data, the members of this cluster may be restraining themselves from using it to a large extent at the inter-organisational level. So, in general, this is the second highly matured cluster. Moreover, the obvious difference between cluster 1 and 4 is the strategic alignment of utilising BDA capabilities. While the members of cluster 1 are using it at the inter-organisational level, members of cluster 4 are strategically using it only within their organisation. This group is certainly more mature and doing better in terms of using data and information for data-driven decision making than the cluster 2 and 3. Based on the characteristics of this cluster, members of this group are called ‘enthusiastic users’- technologically advanced but they have less external focus in terms of BDA usage. Moreover, from RBV perspective, these organisations do possess some unique capabilities that can significantly improve firm performance and create a sustainable competitive advantage. However, in this research, it is identified from the path analysis that organisations should develop and create SCA capabilities and the ability to maintain data and information quality to realise performance improvement. Thus, the members of these organisations can be regarded as those that get a relatively 'temporary competitive advantage'.

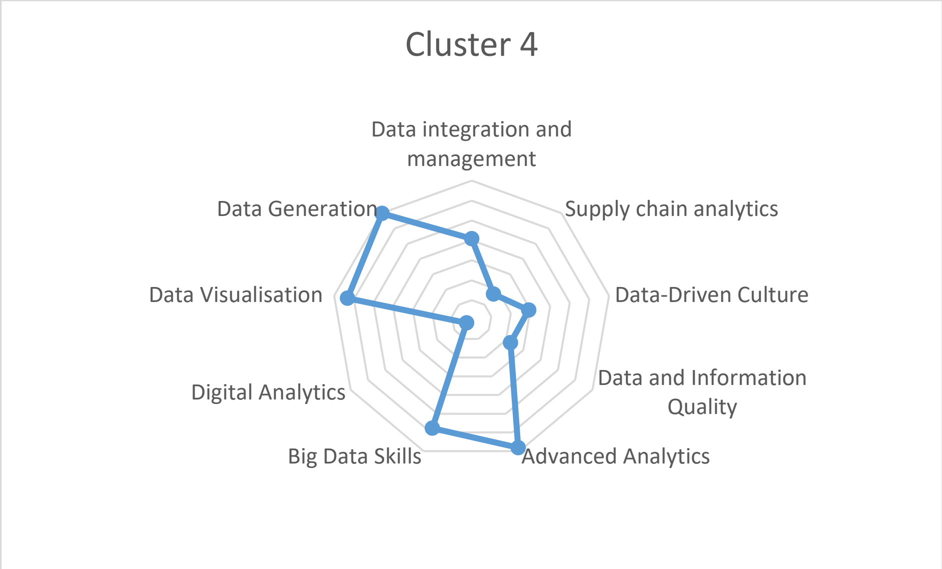
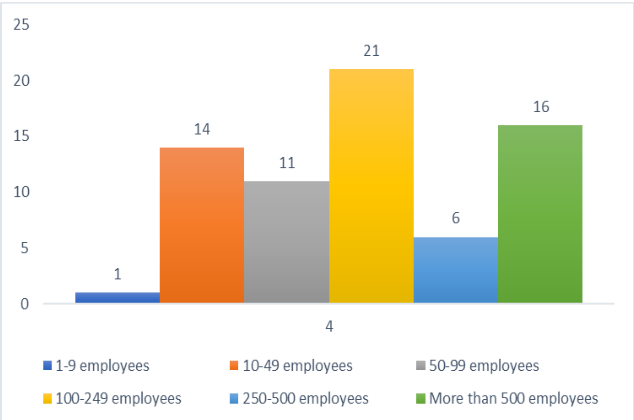


Figure 6.15 Spider plot of cluster 4 Vs BDA capabilities



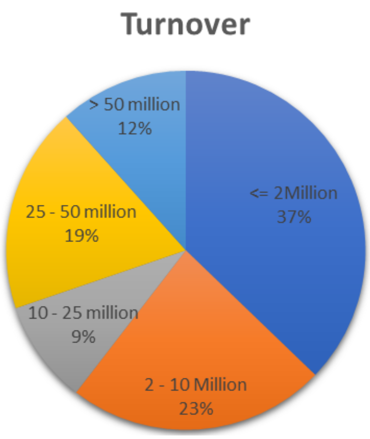


Figure 6.16 a) Cluster 4 Vs Number of employees b) Cluster 4 Vs Annual turnover

In summary, it is evident that some organisations are highly mature in terms of BDA capabilities and some are not. Members of cluster 1 and cluster 4 can be considered as 'mature', but cluster 2 and 3 are relatively 'immature'. The presence of organisations at the two extreme poles, indicates the reality of the digital divide between them. Based on the maturity of BDA capabilities, cluster 1 can be considered as highly 'data rich' and 'information rich', but cluster 3 is relatively 'data poor' and 'information poor'. Moreover, the majority of organisations in cluster 1 are medium to large firms with more number of employees and high annual turnover compared to organisations in cluster 3 who are mostly small firms.

6.5 Managerial implications

Based on the findings of this study, this section presents: 1 the issues and challenges of adopting and practising BDA, and 2. the managerial implications.

6. 5.1 Issues and challenges of adopting and practising BDA

6.5.1.1 Organisational challenges

a) *Time-consuming:* A predictive analytics initiative is time-consuming and includes various stages of developing, testing and adapting it to different contexts (Blackburn et al. 2015). Bringing together experts from various functions with varied mindsets will be a challenging task. In complex systems like supply chains, BDA implementation needs consistent support from top management and key stakeholders, as it might take 12-18 months to see the results. Getting access to data (owned by different departments of the organisation), combining, validating, and cleansing the data will be tedious which requires exhaustive commitment from the project management team.

b) *Insufficient resources:* The data and analytics resource capabilities vary across firms in a supply chain network. Supply chain partners' lack of IT resources and the capability to share data and information in real-time will cause discrepancies. As stated by Dutta and Bose (2015), collaboration and cross-functional team formation between various stakeholders within an organisation should be a priority for implementation of big data. However, while forming an inter-organisational cross-functional team, the challenges that can be anticipated are competition within the supply chain network, principle-agent conflicts, incentives arrangements, data sharing policies, etc. Data-driven culture, which is one of the key BDA capabilities, and fact-based management, should be encouraged across the supply chain network as a strategy for effective utilisation of BDA and to create business value. Leadership also plays a significant role in the successful implementation of BDA systems for SCM (Seah et al. 2010).

c) *Privacy and security concerns:* Big Data possess several concerns such as privacy, security, the unethical use of Big Data and processing data ineffectively (Hu et al. 2014), which would lead to biased findings (Tien 2012). Supply chain professionals have raised concerns about privacy and data security and argued that out-dated regulations are one of the major obstacles in data sharing, especially consumer data (Richey Jr et al. 2016). Privacy, security and data laws could be of serious concern for multinational supply chains, obligated to abide by the laws of different countries while sharing data across supply chains (Alfaro et al. 2015). However, these challenges could be overcome by employing an effective data governance initiative within the process of data integration and management.

d) *Behavioural issues:* From the behavioural perspective, the use of real-time data and information could be challenging because decision makers may excessively react to even small changes in the physical world which would worsen the "bullwhip effect" and increase supply chain risk and cost of inventory (Tachizawa et al. 2015). Tachizawa et al (2015) argued that concerning Big Data there is a risk of identifying many statistically significant but irrelevant correlations that do not have a causal linkage. Big Data focuses predominantly on correlation but not on causation, which necessitates human inferences to solve big problems. Nevertheless, Google has proposed a heuristic approach to solve the correlation and causality problems (Radke and Tseng 2015).

e) *Issues with Return On Investment (ROI)*: Unclear benefits and ambiguity on ROI make stakeholders apprehensive about implementing BDA (Tee et al. 2007; Richey Jr et al. 2016). Achieving financial benefits from BDA is challenging too as it depends mostly on the "downstream" employees who perform the task (Davenport et al. 2001). For instance, analytics can help segment the market based on available data, but it is the sales/marketing team who has to believe in the data-driven insights and treat customers based on the segment types to make real change. Here, data-driven culture and the employees' absorptive capacity at the individual level plays a crucial role in absorbing and assimilating the knowledge.

f) *Lack of skills:* Schoenherr and Speier-Pero (2015) identified potential barriers to using predictive analytics in SCM through a survey. The primary barriers are inexperienced employees, time constraints, lack of integration, lack of appropriate predictive analytics solution, and issues with change management. Moreover, it is found out that professionals who plan to use analytics in the future and who are currently not using it, have considered the lack of data and inability to identify suitable data as a prominent barrier. This relates to the situation of ‘Data Poor and Information Poor’. Waller and Fawcett (2013 b) argued that a data scientist requires a combination of both analytical skills and domain knowledge, and it is difficult to find such a combination as someone who is good in analytical skills may not be interested in learning domain knowledge. A recent study has also confirmed that lack of experts in BDA is a serious issue among supply chain professional (Richey Jr et al. 2016).

6.5.1.2 Technical challenges

a) *Data scalability:* Richey Jr et al. (2016) identify data scalability as a major technical issue in BDA adoption. Organisations have to dump their data after a particular period of time so as to store newly generated data. Replacing relational databases which are limited regarding scalability with more advanced infrastructure such as Hadoop distributed databases, distributed file systems, parallel computing and cloud computing capability could be considered to tackle scalability issues. NoSQL database which has a high level of scalability is a better choice to deal with unstructured data generated from IoT data sources (Kang et al. 2016). However, leveraging cloud-computing capability to store BD could incur more financial burden to organisations as with increased generation of BD cloud storage utilisation cost will also eventually increase. So as to avoid this, organisations could adopt strategies to optimise the data collection process and reduce unwanted data generation right from the source (Rehman et al. 2016).

b) *Data Quality*: Supply chain managers rely on data-driven insights for various reasons such as to gain visibility, collaboration, process control, monitoring, optimisation, etc. and, ultimately aim to obtain competitive advantage (Hazen et al. 2014; Davenport 2006). However, there are quality issues associated with the process of data production, which is often compared with the product manufacturing process (Wang 1998; Wang et al. 1995; Wang and Strong 1996; Hazen et al. 2014). Hazen et al. (2014) stated that poor data quality would hinder the data analytics activities and affect management decisions. Unlike a physical product, data is intangible in nature and measuring data quality is a multidimensional problem (Hazen et al. 2014). Concerns about the trustworthiness of the social media and web scraped data are also raised (Tan et al. 2015). The efficiency of the physical flow of material can be determined by the infrastructure quality (such as transportation system, ports, technology, etc.) (Bagchi et al. 2014). Therefore, for effective flow of information, it is vital to have advanced data infrastructures and best practices of data management using tools and techniques such as Hadoop, MapReduce, and statistical process control.

c) *Lack of techniques and procedures:* Lack of quality data is not the only problem; there is an incapability of techniques to exploit the data deluge properly. For instance, in the case of demand forecasting techniques, significant attention is given solely to endogenous time-serious variables for demand forecasting, and there is a lack of consideration of exogenous variables and information sources (Meixell and Wu 2001). This evidently has inferences to develop better data management capabilities and methodology for forecasting demand to enhance supply chain operations. Also, there is a difficulty in considering expert judgement as a covariate in forecasting models. Further, ordinal scales are used to measure opinions and experts' judgement whereas, in practice, the available modelling approach of independent and dependent variables are intended for continuous or natural data as input. As discussed in Blackburn et al. (2015), very few studies like Fildes et al. (2009) have reflected on using expert judgement to increase the accuracy of supply chain forecasting.

6.5.2. Implications for best practices of BDA

a) Firstly, for developing data generation capabilities, companies should start building their data infrastructure. There are many organisations which do not generate enormous volume of data by leveraging devices such as RFID and IoT. These organisations could start sensing their everyday business environment by deploying these novel technologies. On the other hand, organisations who are data rich should periodically audit their data sources such as RFID and sensors, to ensure no irrelevant data is being generated due to the malfunctioning of these devices. Moreover, establishing metrics based on business rules, defining appropriate variables to measure, and deploying data reduction strategies could allow organisations to reduce the amount of data being generated at source (Rehman et al. 2016). The unstructured data (e.g. customer reviews) can be streamlined to generate less messy data. These approaches will certainly improve data quality and reduce infrastructure cost. When less big data is generated, only good quality data will be created, thereby minimising data logistics efforts and storage costs.

b) Secondly, developing capabilities to integrate data across supply chains is associated with the practice of information sharing. There should be standard policies and principles established to standardise the process of information sharing (what to share and what not to share) with fellow supply chain members to gain mutual benefits and at the same time retain their individual competitive advantage. As we discussed before, cloud computing can be treated as a complementary resource which can aid information sharing maintaining security and access control.

c) Thirdly, for developing analytics capabilities, companies who have a low level of analytics capabilities could start from developing basic analytics (i.e. descriptive) and moving incrementally towards leveraging advanced analytics such as predictive and prescriptive. This research suggests that predictive analytics is mostly used compared to other forms of analytics, but the analytics models once deployed should be validated periodically to ensure performance. Moreover, several factors could facilitate Big data practice: (i) Clearly defining the measurable outcomes, (ii) Effectively addressing the issues, (iii) Employing a flexible model that enables stability and improves decision making, (iv) Utilising appropriate and most suitable data for the purpose (Radke and Tseng 2015).

d) Fourthly, from the perspective of data-driven culture, the success of predictive analytics depends on the degree of integration into business processes and its acceptance by management as a decision-making tool (Blackburn et al. 2015). Blackburn et al. (2015) suggested that the inclusion of experts' opinion into analytics and better communication of insights from analytics to management are the key ways of achieving integration and acceptance of analytics-driven decision making. Similarly, for successful implementation of Big Data analytics, it is critical to clearly define organisation-specific business requirements and performance measures, create a vision for utilising analytics using maturity models and defining roles based on analytical skills (Aho, 2015). The statistical understanding and critical thinking ability of decision makers can play a significant role in data-driven decision making (Radke and Tseng 2015;

Markham et al. 2015). Figure 6.17, illustrates the implications for organisations with the varying level of BDA maturity.

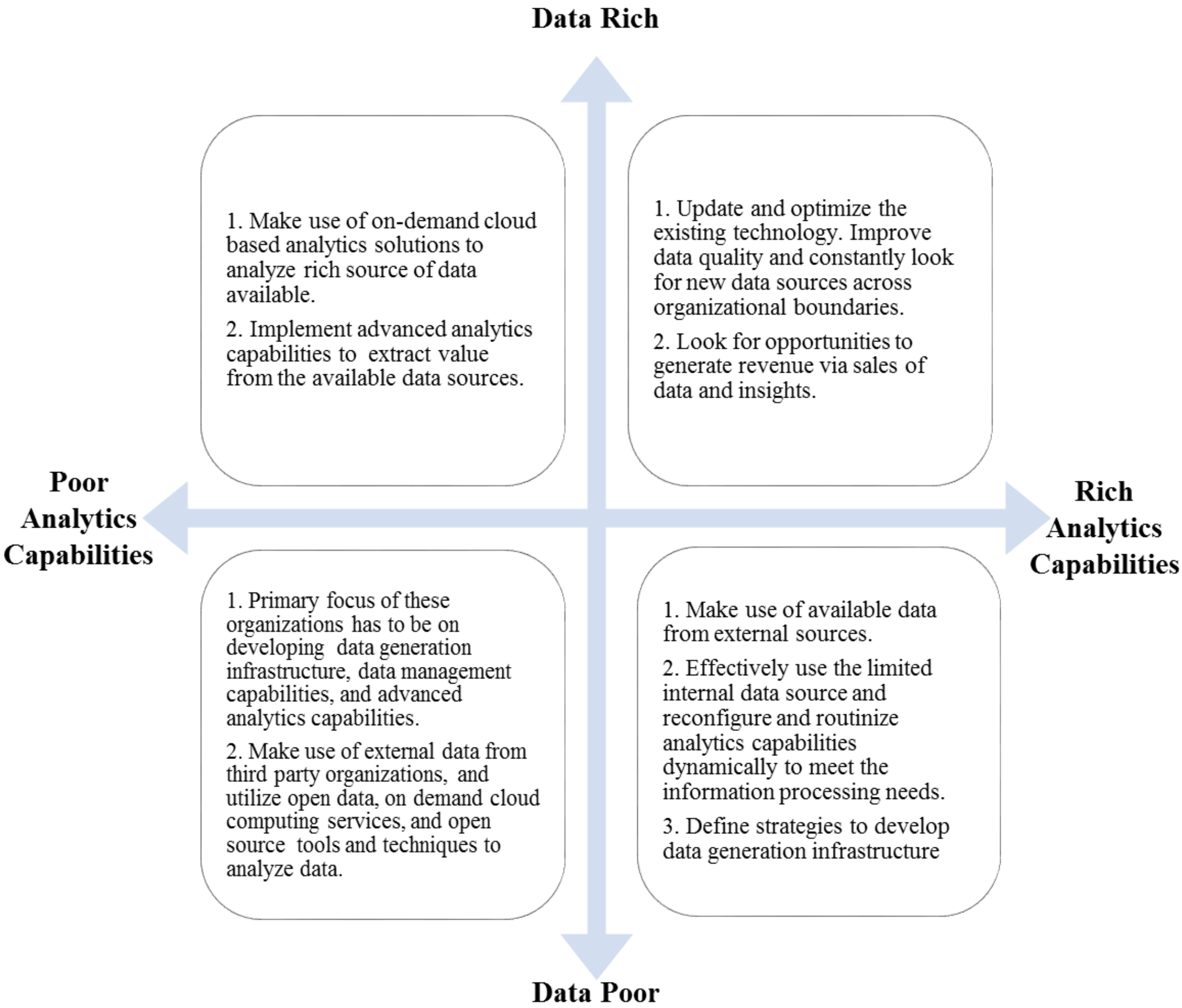


Figure 6.17 Implications for adoption and practice of BDA

Chapter 7 Conclusion

7.1 Chapter overview

This chapter summarises the key outcomes of the research and as well presents the major contributions of it. These are followed by the limitations of this study and the recommendations for future research.

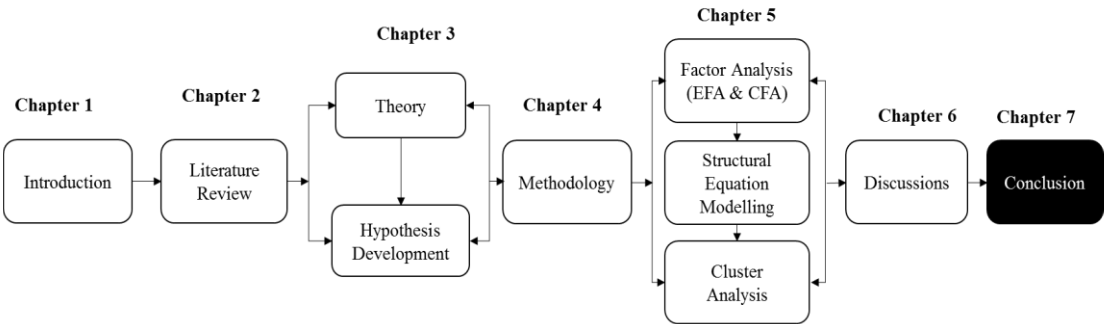


Figure 7.1 Position of the chapter in this thesis

7.2 Summary of the research findings

In summary, BDA has the potential to outperform and transform traditional decision-making practices. This research began with two Systematic Literature Reviews (SLR). In the first SLR, the study reviewed 82 academic papers on Big Data and SCM and 13 maturity models available in both academic and non-academic domains. In the past, BDA has mostly been explored from the technological perspective to rationalise its economic benefits, but the SLR in this research emphasised the necessity to delineate key BDA capabilities that can be used to extract value from big data. The structured approach used for the literature review has revealed existing contributions of BDA and SCM research. Findings from the SLR suggested that BDA could be beneficial if organisations can develop the right capabilities to effectively use the big data. Unlike prior research, the conceptualisation of BDA capabilities in this research is holistic and data-driven. First, it addressed the importance of understanding the genesis of big data generation in the manufacturing supply chain. Second, it suggests the significance of integrating and standardising data from heterogeneous sources to offer more coherent data sets suitable for the analytics systems. Third, different types of analytics and the importance of assimilating the findings into the business process are addressed. In comparison with descriptive and predictive analytics, prescriptive analytics require a negligible amount of human intervention (Puget 2015), which will revolutionise the decision-making process. In the second SLR, the study systematically reviewed 64 academic paper related to

ACAP. There are several important conclusions from the SLR of ACAP. The four research themes identified in this research (technology-driven, relationship-specific, knowledge-specific and ambidextrous absorptive capacity), improve the overall understanding of the concept and how it is applied in the supply chain context.

Although few studies have found significant benefits of BDA, managers still face difficulties extracting the full potential from it (Richey Jr et al. 2016). Using a quantitative approach, the practice of BDA in the UK manufacturing sector is explored to gain an understanding of how BDA impacts firm performance. It is found that organisations can augment the benefits of BDA if they can focus their efforts on developing higher-order capabilities such as absorptive capacity, data and information quality and supply chain analytics capabilities. Using the structural equation modelling technique, the survey data from the UK manufacturing sector confirms these findings. The study also suggests that BDA increase firm performance, by enhancing higher-order capabilities, which then improves operational and innovation performance. The empirical data analysis of this research showed that manufacturing industries which produce several commodities for Business to Business (B2B) or Business to Consumers (B2C) could benefit from the practice of BDA. However, to achieve a sustainable competitive advantage, firms must develop higher-order capabilities such as absorptive capacity, data and information quality and supply chain analytics capabilities. In terms of the digital divide, the study has explored the phenomenon and has found sufficient evidence to argue that there is a divide between small and large UK manufacturing organisations. The conceptual framework presented in this research is holistic and can provide a detailed roadmap for organisations planning to implement BDA in the future.

This study makes significant contributions to both theory and practice. The conceptualisation of BDA capabilities would help academic researchers to embark on new empirical research in this domain. It contributes to the on-going debate of BDA in SCM context and supports a comprehensive understanding of this evolving technology from the systematic literature review and conceptualisation of key capabilities. A holistic conceptual framework of BDA capabilities is developed to describe the stages of BDA assimilation. Moreover, this thesis will guide practitioners to realise their current state of BDA maturity and build a roadmap to develop BDA capabilities keeping in mind the potential challenges associated with the assimilation process. Further, the research also discussed the implications for best practices of BDA in the supply chain. However, there are some limitations of this study and they are addressed in the following section which leads to the identification of future research opportunities.

7.3 Research Contributions

The research process and the resultant outcomes of the present study make significant knowledge contribution to theory, policy, and practice. These contributions are discussed in the following sections.

7.3.1 Contribution to the theory

The desk research phase of this work has made significant contributions in three aspects. First, the aim is to summarise the key components of BDA capabilities into a single conceptual framework (Figure 2.12). This aim is successfully achieved by integrating theoretical knowledge from the works of leading journals in this field. Further, this research has operationalised the BDA capabilities maturity construct. A systematic literature review and a rigorous statistical procedure have supported the establishment of the underlying dimensions of the construct. The data analysis confirms the conceptualisation and operationalisation of Big Data Analytics capabilities maturity as a second-order construct. As a result of SLR and quantitative analysis, this research has revealed seven key dimensions of intra-organisational BDA maturity: data generation capability, data integration and management capability, advanced analytics capability, digital analytics capability, data visualisation capability, data-driven culture and big data skills (Figure 7.2). These capabilities can be grouped into three categories, namely tangible, intangible and human assets. The identification of key conceptual elements of BDA composed of five tangible assets (data generation, data integration and management, advanced analytics, digital analytics, and data visualisation), a human asset (Big Data skills) and an intangible asset (data-driven culture) as the dimensions of the construct is an important knowledge contribution to the literature on BDA and SCM. As suggested by the theory of dynamic capabilities, it can be noted that conceptualising integrated BDA capabilities (BDA maturity) is less likely to be mimicked by competitors as it is a combination of lower-order tangible and complementary human and intangible resources. These findings suggest that manufacturing organisations can exploit these capabilities to enhance operational and innovation capabilities. This research also developed a measurement scale for the BDA maturity construct by combining scales from various sources and validating it empirically. In future, BDA researchers may adopt this verified scale to conduct researcher works and develop it further. The novel method of grouping various elements of BDA practice based on RBV contributes to the theory building by equipping the academic community with an all-inclusive BDA capabilities framework.

This research makes a significant contribution in terms of comprehensively reviewing existing intellectual contributions in the domain of ACAP and SCM. It provides an overview of supply chain researchers' understanding of the concept and how it is used. In the descriptive analysis, the increasing trend of publication on this topic is clearly depicted. In the content analysis, the non-cohesiveness of the conceptual understanding of supply chain researchers is witnessed, and the application of absorptive capacity at the inter-organisational level is found to be high compared to other levels. Through thematic analysis, some important research themes and key propositions are derived.

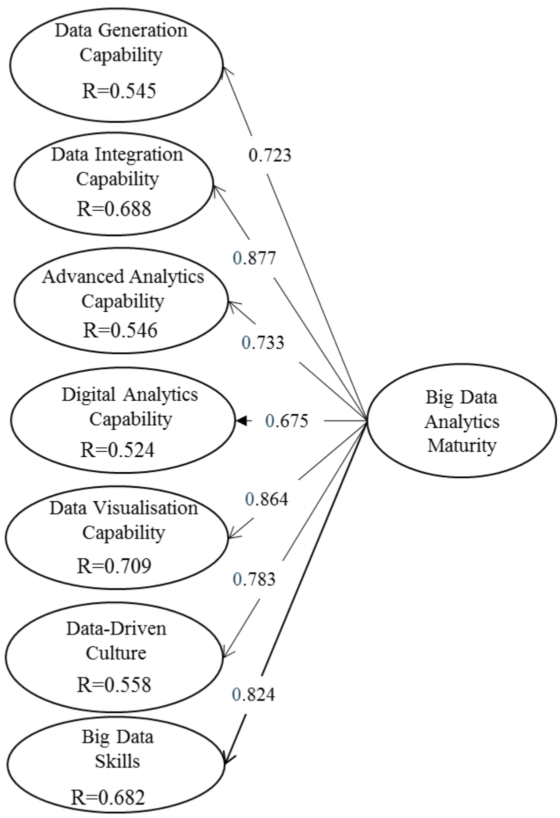


Figure 7.2 Conceptualisation and Operationalisation of key dimensions of BDA Maturity

In addition, this research has applied SLR techniques which highlights the value of this research as SLR is regarded as a scientific technique (Tranfield et al. 2003). The SLR techniques used in this research have great potential for the progress of research practice in management discipline. Thus, this aspect of the research depicts one of the main motivations of this thesis, aiming to develop the management discipline by promoting the application of SLR to increase scientific rigour.

Further, the empirical results confirming that BDA capabilities predict product quality, cost, flexibility, time and innovation performance is an important theoretical contribution. This study integrates the theoretical view of RBV and dynamic capabilities view to advace the understanding of underlying mechanism through which firms achieve competitive advantage. This study contributes to these theories by empirically verifying that organisational learning, quality of information and inter-organisational data capabilities are the underlying reason for creating sustainable competitive advantage. Drawing upon the dynamic capabilities view, this study has proved that organisation that process data and information dynamically can attain significant competitive advantage in the global market place. In a global market, manufacturing firms need to win orders from both domestic and global customers and the firm performance dimensions measured in this research are important criteria that can win orders. The current market environment is dynamic and surrounded by uncertainties due to unpredictable consumer demands and requests. Thus, the conversion of raw data into valuable knowledge by utilisation of dynamic BDA capabilities is an important theoretical contribution.Moreover, this study also contributes to the view of hierarchy of capabilities by determing that capabilities are hierarchical in nature and developing higher-order dynamic, organisational and operational capabilities may lead to improved competitive advantage through the generation of economic rents. By confirming the significant positive effect of lower-order capabilities on higher-order capabilities such as absorptive capacity this study makes a significant contribution for building of dynamic capabilities view theory from hierarchical perspective. Subsequently, this research supports the view that developing a hierarchy of capabilities is a basis for creating a sustained competitive advantage.

Additionally, there is a broad opinion in the literature that SMEs completely lack competence in Big Data Analytics. The findings of this research suggest that this belief about SMEs' maturity level may not be completely true. This study has shown that SMEs, especially medium-sized manufacturing firms, do possess some BDA capabilities. However, small firms do lack BDA capabilities or any form of data-driven decision making. Further, prior BDA research has mainly focused on large firms. However, this study has increased the understanding of the current state of SMEs in the UK manufacturing sector and their BDA capabilities, and thereby contributing to the literature on SMEs.

In summary, this study's contribution to theory can enhance our understanding of the latent factors that form BDA capabilities maturity, and their relationship with higher-order capabilities such as absorptive capacity, data and information quality, and supply chain analytics capabilities. This study also provides knowledge on the nature of the formation of lower-order and higher-order capabilities and how these act as significant antecedents of operational and innovation performance.

7.3.2 Contribution to practice

The findings of this study make several contributions to practice. The development of a measurement scale for BDA capabilities maturity is an important contribution for practice as well. Using the maturity model and the validated scales to measure BDA capabilities, a practitioner could assess the current level of a firm's BDA capabilities and determine the areas for improvement. The study contributes to the practices by offering knowledge on key capabilities that influence firm performance. Also, the BDA capabilities framework developed in this study (Figure 2.12) and the implications for the best practice of BDA (section 6.5) could facilitate firms to extract full value from their investment. Moreover, for firms interested in developing BDA capabilities, this research offers guidance on where to start from. This study has recognised the prominence of higher-order capabilities by evaluating its impact on operational and innovation performance.

Manufacturing companies have changed their strategy in response to the changing market demands and have become more customer-centric (Aho 2015). Manufacturing companies consider embedding services into products as a differentiation strategy. The transformation of manufacturing firms from the traditional product offering to servitization involves developing new capabilities such as BDA capabilities. BDA maturity model would facilitate organisations to create a vision for utilizing analytics and adding value to processes by comparing against best practices and external benchmarks (Aho 2015).

BDA considerably impacts on business innovations. It is imperative for manufacturing firms to make use of BDA capabilities in innovation processes such as new product development as they can offer vital information for idea generation. Managers aiming to improve the efficiency and effectiveness of innovation processes should enhance their employees' competence to effectively leverage BDA tools. To strengthen the effect of BDA capabilities, firms should strategically align the BDA tools and structural and cultural conditions of the organisation with the requirements of their innovation processes.

Moreover, Big data skills are highly desired to increase the effectiveness of BDA practice. Studies show that there is a significant skills shortage across the world (Debortoli et al. 2014). Some technology vendors such as IBM are partnering with a large number of universities around the world to narrows the skills gap (IBM 2013). Similarly, managers of manufacturing firms should collaborate with universities to recruit qualified BDA personnel such as data scientists and data engineers to encourage the best practice of BDA.

7.3.3 Contribution to policy

The insights offered in this research give rise to policy recommendations that could potentially help the UK manufacturing SMEs. First, the conceptualisation of dimensions of BDA capabilities and the measurement scales developed in this research could be utilised by the members of government agencies to evaluate the current level of BDA capabilities of manufacturing SMEs. This kind of quantitative evaluation of SMEs could help them to determine the strengths, weaknesses, and challenges of SMEs relating to the adoption and practice of big data analytics. This can enable the government agencies to determine the appropriate support mechanism required to enable SMEs to develop the BDA capabilities that they are deficient in.

The knowledge from this research could also contribute to the development of policies relating to SMEs growth. A policy could be established to encourage key stakeholders of SMEs to adopt the data-driven culture, data-driven decision-making practices, and systems thinking as a basis for developing highly advanced BDA capabilities. This could possibly be achieved by raising awareness of owner-managers of SMEs about the business value rooted in the practice of BDA. By creating this awareness, the adoption of data-driven decision-making capabilities of SMEs can be enhanced - providing a competitive advantage.

The finding that BDA capabilities can significantly promote innovation has a policy implication. Considering the paradigm shift from an industrial economy to the innovation economy, a government policy could be developed to support High-Value Manufacturing (HVM) firms driven by innovation. HVM is the application of cutting-edge technology and technical know-how to create products, processes, and services that have great potential to bring sustainable economic growth to the UK. In fact, the innovate UK initiative of the UK government considers the promotion of HVM as a sustainable strategy for economic value creation (UK Government 2014). High-value manufacturing is characterised by the presence of high-level of R&D intensity, and to innovate, knowledge from a wide range of sources must be applied. By providing a support mechanism to implement BDA, government agencies could encourage SMEs and large organisations who operate in HVM space to create effective and efficient innovation systems.

BDA is found to be an important facet of manufacturing management. However, the application of BDA as a strategic resource is found to be scarce among the small organisations investigated in this study. The underlying cause of this could be due to cost and lack of awareness. As a part of government support, vendors could be encouraged to develop BDA systems catering to the requirements of SMEs and distribute them at a subsidised cost. Such a policy could be a catalyst for smooth transitioning of SMEs from traditional heuristics-based decision-making to data-driven decision-making. Moreover, data analytics is becoming more affordable. For instance, the storage cost of data is declining at approximately 40% per year, and the development of solid-state drives (SSDs) will further intensify the reduction of storage costs (OECD 2014). The use of open source solutions such as Hadoop and CouchDB are at the forefront of today's big data processing, and SMEs could benefit from these cost reductions in technology. Consequently, BDA will become more affordable for SMEs and start-ups, and their adoption will be intensified as the data volume increases and if the necessity of data-driven insights for better competitiveness is recognised by the stakeholders.

Finally, the policymakers should recognise that the disparity in the adoption of BDA could favour the condition of information asymmetry between SMEs and large organisations and thus bring in power shifts. This leads to the formation of a new kind of digital divide and could weaken economic resilience (OECD 2015). The findings of this study could be beneficial to the UK government agencies such as Department for Business, Energy and industrial strategy (BEIS) to develop strategies for the growth the UK manufacturing companies and provide an opportunity to effectively compete in the global market.

7.4 Limitations of this research

Despite the significant contributions of this study, there are a few limitations that deserve attention.

With regard to the data collection, the data is mainly from a single sector (manufacturing) in one specific country. The study did not include organisations from around the world, the results may be limited to the specific context of the UK and lack generalisability (Bryman and Bell 2011). However, as a critical realist, it can be argued that this research mainly seeks to "generalize, not about populations, but about theoretical propositions" (Edwards et al. 2014, p.18). Moreover, with regard to the sample size, certainly a larger sample size can increase confidence towards generalising the findings in the specific context. However, in this research, the sample size is comparatively larger than most of the survey-based research papers in supply chain management discipline (de Beuckelaer and Wagner 2012), increasing the significance of the data collected. Further, the sample size should be seen taking into consideration the context of this research, which is limited to only the UK manufacturing sector. In addition, the research involves receiving responses from senior executives from the manufacturing organisations. Therefore, the sample size can be considered adequate to offer insights into the impact of BDA on the UK manufacturing organisations.

In terms of methodology, even though this research adopted the rigorous quantitative approach, a qualitative case study approach would have allowed the researcher to capture an in-depth understanding of the phenomenon (Yin 2003). Further, the methodological approach adopted in this study captured the truth only at one point in time (Bryman and Bell 2011). The dynamics of the impact of BDA on the performance outcomes are not fully captured and the results of this study are limited in this regard. Perhaps adoption of a qualitative interview technique would have aided the researcher to explore and identify unknown variables that can influence the relationship between BDA maturity and firm performance.

In terms of data analysis, this study has conceptualised BDA as a second-order reflective construct, but previous studies have demonstrated the rationale for conceptualising technology related dimensions as a second-order formative construct (Mackenzie et al. 2011; Treiblmaier et al. 2011; Anon 2008; Shin and Kim 2011). There is a growing debate on the use of formative versus reflective constructs in operations and information systems research (Diamantopoulos and Siguaw 2006). Scholars are yet to arrive at a clear consensus regarding the benefits of one approach over the other. Moreover, analysing the structural relationship between formative constructs require the use of Partial Least Square Structural Equation Modelling (PLS-SEM). But, the use of PLS-SEM in operations research is criticised by researchers in the discipline (Rönkkö et al. 2016). On the other hand, studies that compared both PLS-SEM and Covariance Based SEM (CB-SEM) have suggested to use PLS-SEM as it is advantageous even in the case of small sample size (Astrachan et al. 2014) Consequently, using the same dataset, it would be interesting to observe any change in the nature of the relationship between the constructs of the research model if BDA maturity is conceptualised as a formative construct.

Moreover, while a systematic and structured literature review is conducted, it is worth recognising the concerns associated with it. It should be noted that the citation count information extracted from the Scopus database differs from other databases such as WoS or Google scholar. The selection of journal papers is restricted by the use of specific keywords -other keywords and databases could have resulted in additional papers. Even though we relied on specific keywords and databases relevant to this research, efforts were made to validate, avoid errors and ensure that all relevant papers are included. Further, the synthesis of the conceptualisation does involve authors' interpretations of selected journal papers and their previous expertise in BDA domain. However, to ensure reliability the process of selecting journal papers, coding and conceptualisation are verified independently by peer researchers. Moreover, despite including maturity models from practice in the conceptualisation process, most of the papers reviewed are from academic sources.

Nonetheless, it can be argued that the limitations discussed above can be considered as low in view of the significant contributions of this research. Moreover, the scope of this research presented in this chapter 1 highlights how the findings of this research can be generalised to the specific context of the UK manufacturing sector. However, several suggestions for the future research are identified based on the research findings and from the knowledge gained from the execution of the research process, which are discussed in the following section.

7.5 Directions for future research

i. With the growing importance of data-driven decision making in supply chain management, poor data quality would negatively impact supply chain performance. Data quality should be monitored, measured and controlled. In the future, a systematic investigation of various techniques used to tackle the data quality problem in the context of inter-organisational supply chain networks should be investigated.

ii. In-memory analytics is advantageous for SCM as it enables real-time decision-making. While previous case studies (vom Brocke et al. 2014; Piller and Hagedorn 2011) explained the potential benefits of in-memory analytics, there is a need for comprehensive investigation of the real-world use case of in-memory analytics in SCM.

iii. BDA influences a paradigm shift from heuristics to data-driven decision making. The reasoning and assimilation of data-driven insights into the process of decision-making and the behavioural issues associated with it have to be investigated. Moreover, the phenomena of resistance to change from heuristics to data-driven decision making should be explored in different organisational contexts.

iv. There are various examples of successful implementation of Business analytics to optimise supply chain and improve performance (Chae et a. 2014). There are various anecdotal evidence such as company case studies, whitepapers, etc. justifying the potential impact of BDA on firm performance. However, there is some evidence of failure as well. Oliveira et al. (2012) and Sherer (2005) discussed the case of Cisco’s Business Analytics failure. However, some questions remain unanswered, such as, why and under what circumstances do organisations’ BDA initiatives fail? There is a need for in-depth case studies to understand the organisational context under which organisations achieve success or failure from BDA investment.

v. BDA implementation is a change-intensive process and organisations are vulnerable to disruption. Systems design should keep its pace in addressing these changing needs to achieve competitive advantage. Empirical studies are required to describe and understand the mechanisms and dynamics of how BDA implementation influences change and redesign of the system.

vi. With respect to the digital divide, this research is limited by scope and considered only the divide between manufacturing companies. However, future studies could investigate the current state of the retail and other services sector in adopting and practising BDA and compare it with the findings of the manufacturing sector. The nature of data generated and the analysis techniques used by manufacturing and retail industries could vary significantly. By exploring and comparing the maturity level of the manufacturing and retail sector, new insights about BDA practice could be revealed.

vii. From the cluster analysis, it is identified that although some companies lack BDA capabilities at the intra-organisational level, they practice it at the supply chain level to some extent. This indicates there may be an influence of pressure from the external environment to adopt and practice BDA at the supply chain level. According to DiMaggio and Powell (1983), organisations under similar environmental conditions are forced to resemble each other, a process called 'isomorphism'. The theory suggests that institutional isomorphism occurs through coercive, mimetic and normative mechanisms. The choices of the business decision are highly influenced by internal and external norms (Tate et al. 2014). Organisations can be categorised as either reactive adopters or late adopters depending upon their pattern of isomorphism (Tate et al. 2014). In future, the institutional theory can be used to investigate both the types of adoption scenarios and the influence of coercive, mimetic and normative pressures from the supply chain leader in the adoption of BDA as shown in Figure 7.3.

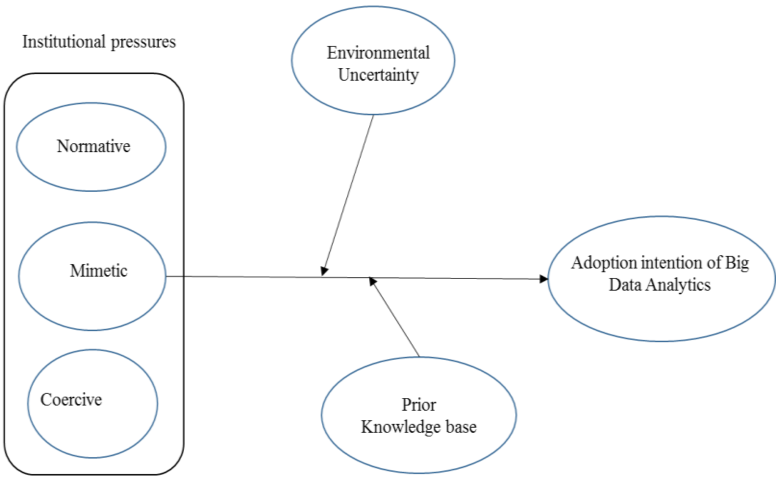


Figure 7.3 Conceptual framework of the impact of institutional pressure on the adoption of BDA

References

Abareshi, A., & Molla, A. (2013). Greening logistics and its impact on environmental performance: an absorptive capacity perspective. *International Journal of Logistics Research and Applications*, *16*(3), 209–226.

Abdullah, N., Ismail, S. A., Sophiayati, S., & Sam, S. M. (2015). Data quality in big data: A review. *International Journal of Advances in Soft Computing and Its Applications*, *7*(3), 16–27.

Acito, F., & Khatri, V. (2014). Business analytics: Why now and what next? *Business Horizons*, *57*(5), 565–570. http://doi.org/10.1016/j.bushor.2014.06.001

Adams, R., Bessant, J., & Phelps, R. (2006). Innovation management measurement: A review. *International Journal of Management Reviews*, *8*(1), 21–47.

Ağan, Y., Kuzey, C., Acar, M. F., & Açıkgöz, A. (2014). The relationships between corporate social responsibility, environmental supplier development, and firm performance. *Journal of Cleaner Production*, *112*, 1872–1881.

Aghion, P., & Jaravel, X. (2015). Knowledge spillovers, innovation and growth. *The Economic Journal Economic Journal*, *125*(583), 533–573

Aho, A. M. (2015). *Product Data Analytics Service model for Manufacturing company*. (L. Uden, M. Heričko, & I.-H. Ting, Eds.), *Lecture Notes in Business Information Processing* (Vol. 224). Cham: Springer International Publishing

Akaike, H. (1973). Information Theory and an Extension of the Maximum Likelihood Principle. Springer New York. http://doi.org/10.1007/978-1-4612-1694-0_15

Aldridge, A., & Levine, K. (2001). *Surveying the social world : principles and practice in survey research*. Buckingham: Open University press.

Alfaro, L. A., Le, T. M. H., Choi, H. R., & Cho, M. J. (2015). Deployment Model of Big Data for Port Logistics. *Information (Japan)*, *18*(1), 7–16.

Ali, M., & Park, K. (2016). The mediating role of an innovative culture in the relationship between absorptive capacity and technical and non-technical innovation. *Journal of Business Research*, *69*(5), 1669–1675.

Allred, C. R., Fawcett, S. E., Wallin, C., & Magnan, G. M. (2011). A Dynamic Collaboration Capability as a Source of Competitive Advantage. *Decision Sciences*, *42*(1), 129–161.

Almada-Lobo, F. (2016). The Industry 4.0 revolution and the future of Manufacturing Execution Systems (MES). *Journal of Innovation Management*, *3*(4), 16–21.

Alttok, T. (1997). *Performance Analysis of Manufacturing Systems*. (P. Glynn, Ed.). Springer Series in Operations Research.

Ambulkar, S., Blackhurst, J. V., & Cantor, D. E. (2016). Supply chain risk mitigation competency: an individual-level knowledge-based perspective. *International Journal of Production Research*, *54*(5), 1398–1411.

Amit, R., & Schoemaker, P. J. H. (1993). Strategic Assets and Organizational Rent. *Strategic Management Journal*, *14*(1), 33–46.

Amoako-Gyampah, K., & Acquaah, M. (2008). Manufacturing strategy, competitive strategy and firm performance: An empirical study in a developing economy environment. *International Journal of Production Economics*, *111*(2), 575–592.

Amoako-Gyampah, K., & Boye, S. S. (2001). Operations strategy in an emerging economy: The case of the Ghanaian manufacturing industry. *Journal of Operations Management*, *19*(1), 59–79. http://doi.org/10.1016/S0272-6963(00)00046-2

Arnold, V., Benford, T., Hampton, C., & Sutton, S. G. (2010). Competing pressures of risk and absorptive capacity potential on commitment and information sharing in global supply chains. *European Journal of Information Systems*, *19*(2), 134–152.

Arnold, V., Benford, T., Hampton, C., & Sutton, S. G. (2011). Enterprise Risk Management as a Strategic Governance Mechanism in IT-enabled Transnational Supply Chains. *Journal of Information Systems*, *26*(July), 51–76.

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

Arunachalam, D., & Kumar, N. (2018). Benefit-based consumer segmentation and performance evaluation of clustering approaches: An evidence of data-driven decision-making. *Expert Systems with Applications*, 1–24.

Arunachalam, D., Kumar, N., & Kawalek, J. P. (2018). Understanding big data analytics capabilities in supply chain management: Unravelling the issues, challenges and implications for practice. *Transportation Research Part E: Logistics and Transportation Review*, *114*, 416–436. http://doi.org/10.1016/j.tre.2017.04.001

Assunção, M. D., Calheiros, R. N., Bianchi, S., Netto, M. A. S., & Buyya, R. (2015). Big data computing and clouds: Trends and future directions. *Journal of Parallel and Distributed Computing*, *79–80*, 3–15

Astrachan, C. B., Patel, V. K., & Wanzenried, G. (2014). A comparative study of CB-SEM and PLS-SEM for theory development in family firm research. *Journal of Family Business Strategy*, *5*(1), 116–128.

Autry, C. W., Grawe, S. J., Daugherty, P. J., & Richey, R. G. (2010). The effects of technological turbulence and breadth on supply chain technology acceptance and adoption. *Journal of Operations Management*, *28*(6), 522–536.

Ayanso, A., Cho, D. I., & Lertwachara, K. (2010). The digital divide: Global and regional ICT leaders and followers. *Information Technology for Development*, *16*(4), 304–319.

Azadegan, A. (2011). Benefiting from supplier operational innovativeness: The influence of supplier evaluations and absorptive capacity. *Journal of Supply Chain Management*, *47*(2), 49–64.

Azadegan, A., & Dooley, K. J. (2010). Supplier innovativeness, organizational learning styles and manufacturer performance: An empirical assessment. *Journal of Operations Management*, *28*(6), 488–505.

Azadegan, A., Dooley, K. J., Carter, P. L., & Carter, J. R. (2008). Supplier innovativeness and the role of interorganizational learning in enhancing manufacturer capabilities. *Journal of Supply Chain Management*, *44*(4), 14–35.

Babiceanu, R. F., & Seker, R. (2016). Big Data and virtualization for manufacturing cyber-physical systems: A survey of the current status and future outlook. *Computers in Industry*.

Bagchi, P., Lejeune, M. A., & Alam, A. (2014). How supply competency affects FDI decisions: Some insights. *International Journal of Production Economics*, *147*, 239–251.

Bagozzi, R. P., & Yi, Y. (2012). Specification, evaluation, and interpretation of structural equation models. *Journal of the Academy of Marketing Science*, *40*(1), 8–34.

Bagozzi, R. P., Yi, Y., & Phillips, L. W. (1991). Assessing Construct Validity in Organizational Research. *Administrative Science Quarterly*, *36*(3), 421.

Banomyong, R., Varadejsatitwong, P., & Oloruntoba, R. (2017). A systematic review of humanitarian operations, humanitarian logistics and humanitarian supply chain performance literature 2005 to 2016. *Annals of Operations Research*, 1–16.

Barad, M., & Sipper, D. (1988). Flexibility in manufacturing systems: definitions and petri net modelling. *International Journal of Production Research*.

Barlow, M. (2013). *Real-Time Big Data Analytics: Emerging Architecture*. *Journal of Chemical Information and Modeling* (1st ed., Vol. 53). O'Reilly Media, Inc.

Barney, J. (1991). Firm resources and sustained competitive advantage. *Journal of Management*, *17*(1), 99–120.

Barney, J. (2001). The resource-based view of the firm: Ten years after 1991. *Journal of Management*, *27*(6), 625–641

Barney, J. B. (1986). Organizational Culture: Can It Be a Source of Sustained Competitive Advantage? *Academy of Management Review*, *11*(3), 656–665.

Barney, J. B., & Clark, D. N. (2007). *Resource-Based Theory: Creating and Sustaining Competitive Advantage*. *Oxford University Press.*

Barney, J., & Hesterly, W. S. (2012). *Strategic Management and Competitive Advantage: Concepts*. New Jersey: Pearson Education Limited.

Baron, R. M., & Kenny, D. a. (1986). The Moderator-Mediator Variable Distinction in Social The Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations. *Journal of Personality and Social Psychology*, *51*(6), 1173–1182

Barreto, I. (2009). Dynamic Capabilities: A Review of Past Research and an Agenda for the Future. *Journal of Management*, *36*(1), 256–280.

Bartholomew, D. J., Knott, M., & Moustaki, I. (2011). *Latent Variable Models and Factor Analysis. A unified approach*. A John Wiley & Sons, Ltd.

Bartlett, M. S. (1954). A note on the multiplying factors for various X2 approximations. *Journal of the Royal Statistical Society*, *16*(2), 296–298.

Batini, C., & Scannapieco, M. (2006). *Data Quality*. *Springer*. http://doi.org/10.1007/3-540-33173-5

Batini, C., & Scannapieco, M. (2016). *Data and Information Quality Dimensions, Principles and Techniques*. Springer. http://doi.org/10.1007/978-3-319-24106-7_1

Bauer, A., Bauer, A., & J. Browne. (1994). *Shop Floor Control Systems: From design to implementation*. Chapman & Hall.

Beamon, B. M. (1999). Measuring supply chain performance. *International Journal of Operations & Production Management*, *19*(3), 275–292.

Becker, G. S. (1975). *Human capital : a theoretical and empirical analysis, with special reference to education*. Chicago: University of Chicago Press. Retrieved from http://encore.newcastle.edu.au/iii/encore/record/C__Rb1311290__Shuman capital__Orightresult__U__X7?lang=eng&suite=cobalt

Beheregarai Finger, A., B. Flynn, B., & Laureanos Paiva, E. (2014). Anticipation of new technologies: supply chain antecedents and competitive performance. *International Journal of Operations & Production Management*, *34*(6), 807–828.

Belekoukias, I., Garza-Reyes, J.A. and Kumar, V. (2014). The impact of lean methods and tools on the operational performance of manufacturing organisations, *International Journal of Production Research*, 52(18), pp.5346–5366

Bell, J., & Waters, S. (2014). *Doing Your Research Project:A Guide for First-time Researchers* (6th ed.). Berkshire: McGraw-Hill Education.

Bellamy, M. A., Ghosh, S., & Hora, M. (2014). The influence of supply network structure on firm innovation. *Journal of Operations Management*, *32*(6), 357–373.

Bentler, A. (1990). Comparative Fit Indices in Structural Models. *Psychological Bulletin*, *107*(2), 238–246. Retrieved from https://escholarship.org/uc/item/2mk8r49v

Bentler, P. M. (2010). SEM with simplicity and accuracy. *Journal of Consumer Psychology*, *20*(2), 215–220. http://doi.org/10.1016/j.jcps.2010.03.002.SEM

Bentler, P. M., & Kano, Y. (1990). On the Equivalence of Factors and Components. *Multivariate Behavioral Research*, *25*(1), 67–74.

Berghman, L., Matthyssens, P., & Vandenbempt, K. (2012). Value innovation, deliberate learning mechanisms and information from supply chain partners. *Industrial Marketing Management*, *41*(1), 27–39. http://doi.org/10.1016/j.indmarman.2011.11.014

Beyer, M. A., & Laney, D. (2012). The Importance of “Big Data”: A Definition. *Gartner Publications*, *i*(June), 1–9.

Bhakoo, V., & Choi, T. (2013). The iron cage exposed: Institutional pressures and heterogeneity across the healthcare supply chain. *Journal of Operations Management*, *31*(6), 432–449. http://doi.org/10.1016/j.jom.2013.07.016

Bharadwaj, A. S. (2000). A resource-based perspective on information technology capability and firm performance: An empirical investigation. *MIS Quarterly: Management Information Systems*, *24*(1), 169–193. http://doi.org/10.2307/3250983

Bhattacharjya, J., Ellison, A., & Tripathi, S. (2016). An exploration of logistics-related customer service provision on Twitter: The case of e-retailers. *International Journal of Physical Distribution & Logistics Management, 46*(6/7), 659–680.

Bi, Z., & Cochran, D. (2014). Big data analytics with applications. *Journal of Management Analytics*, *1*(4), 249–265. http://doi.org/10.1080/23270012.2014.992985

BIS Department for Business Innovation & Skills. (2010a). *Manufacturing in the UK : An Economic Analysis of the Sector*. Retrieved from http://www.bis.gov.uk/assets/biscore/business-sectors/docs/m/10-1333-manufacturing-in-the-UK-an-economic-analysis-of-the-sector

BIS Department for Business Innovation & Skills. (2010b). *Manufacturing in the UK: Supplementary analysis*. Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/31786/10-1334-manufacturing-in-the-UK-supplementary-analysis.pdf

BIS Department for Business Innovation & Skills. (2015). *Business Population Estimates for the UK and Regions 2015*. Retrieved from https://www.gov.uk/government/statistics/business-population-estimates-2015

Bititci, U. S., Carrie, A. S., & McDevitt, L. (1997). Integrated performance measurement systems: a development guide. *International Journal of Operations & Production*

Management, *17*(5), 522–534. http://doi.org/10.1108/01443579710167230

Blackburn, J. D. (1991). *Time-based competition: the next battleground in American manufacturing*. Irwin Professional Pub.

Blackburn, J. D. (1992). Time-Based Competition: White-Collar Activities. *Business Horizons*, (August), 96–101.

Blackburn, R., Lurz, K., Priese, B., Göb, R., & Darkow, I.-L. (2015). A predictive analytics approach for demand forecasting in the process industry. *International Transactions in Operational Research*, *22*(3), 407–428. http://doi.org/10.1111/itor.12122

Blocher, E. J., Stout, D. E., & Cokins, G. (2010). *Cost Management: a strategic emphasis*. McGraw Hill. http://doi.org/10.1201/b17643-7

Blohm, I., Leimeister, J. M., & Krcmar, H. (2013). Crowdsourcing: How to Benefit from (Too) Many Great Ideas. *MIS Quarterly Executive*, *12*(4), 199–211.

Bock, S. S., & Isik, F. F. (2015). A new two-dimensional performance measure in purchase order sizing. *International Journal of Production Research*, *53*(16), 4951–4962. http://doi.org/10.1080/00207543.2015.1005769

Boddy, D. (2014). *MANAGEMENT: An Introduction*. London, UK: Pearson Education Limited.

Bollen, K. A. (1989). Structural equations with latent variables. *Wiley Series in Probability and Mathematical Statistics*, *8*, 528. http://doi.org/10.1007/s13398-014-0173-7.2

Bollen, K. A., & Pearl, J. (2013). Eight Myths About Causality and Structural Equation Models. In S. L. Morgan (Ed.), *Handbook of Causal Analysis for Social Research* (pp. 301–328). Dordrecht: Springer Netherlands. http://doi.org/10.1007/978-94-007-6094-3_15

Bollen, K. A., & Stinet, R. (1990). Direct and Indirect Effects : Classical and Bootstrap Estimates of Variability. *Sociological Methodology*, *20*, 115–140.

Bonnes, K. (2014). Predictive Analytics for Supply Chains : a Systematic Literature Review. *Proceedings of the BPM Demo Sessions 2014 Co-Located with the 12th International Conference on Business Process Management (BPM 2014)*.

Boone, C. A., Hazen, B. T., Skipper, J. B., & Overstreet, R. E. (2016). A framework for investigating optimization of service parts performance with big data. *Annals of Operations Research*. http://doi.org/10.1007/s10479-016-2314-1

Bowman, C., & Ambrosini, V. (2003). How the Resource-based and the Dynamic Capability Views of the Firm Inform Corporate-level Strategy. *British Journal of Management*, *14*(4), 289–303. http://doi.org/10.1111/j.1467-8551.2003.00380.x

Boyd, D., & Crawford, K. (2012). Critical Questions for Big Data. *Information, Communication & Society*, *15*(5), 662–679.

Brandau, A., & Tolujevs, J. (2013). Modelling and Analysis of Logistical State Data. *Transport and Telecommunication*, *14*(2), 102–115. http://doi.org/10.2478/ttj-2013-0009

Brennan, L. L. (2011). *Operations Management*. Retrieved from http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:No+Title#0

Bridges Jr, C. C. (1966). Hierarchical cluster analysis. *Psychological Reports*, *18*(3), 851–854.

Briggs, A. R., Morrison, M., & Coleman, M. (2012). *Research methods in educational leadership and management* (2nd ed.). London, UK: SAGE puplications.

Brown, R. L. (1997). Assessing specific mediational effects in complex theoretical models. *Structural Equation Modeling*, *4*(2), 142–156.

Bruque-Camara, S., Moyano-Fuentes, J., & Maqueira-Marin, J. M. (2016). Supply chain integration through community cloud: Effects on operational performance. *Journal of Purchasing and Supply Management*, *22*(2), 141–153.

Bruque Camara, S., Moyano Fuentes, J., & Maqueira Marin, J. M. (2015). Cloud computing, Web 2.0, and operational performance The mediating role of supply chain integration. *International Journal Of Logistics Management*, *26*(3), 426–458.

Brusco, M. J., Singh, R., Cradit, J. D., & Steinley, D. (2017). Cluster analysis in empirical OM research: survey and recommendations. *International Journal of Operations & Production Management*, *37*(3), 300–320.

Brusset, X. (2016). Does supply chain visibility enhance agility? *International Journal of Production Economics*, *171*, 46–59. http://doi.org/10.1016/j.ijpe.2015.10.005

Bryman, A. (2012). *Social research methods* (4th ed.). Oxford University Press.

Bryman, A., & Bell, E. (2011). *Business Research Methods*. *Oxford University Press, Oxford, UK.* (3rd ed.). Oxford University Press Inc.

Brynjolfsson, E., & McAfeeErik, A. (2012). Big Data: The Management Revolution. *Harvard Business Review*, *90*(10), 60–68.

Burrell, G., & Morgan, G. (1980). *Sociological Paradigms and Organisational Analysis*. Heinemann Edu:ational Books Ltd. http://doi.org/10.1177/003803858001400219

Busenitz, L. W., & Barney, J. B. (1997). Differences between entrepreneurs and managers in large organizations: Biases and heuristics in strategic decision-making. *Journal of Business Venturing*, *12*(1), 9–30.

Buzzell, R. D. (2004). The PIMS program of strategy research - A retrospective appraisal. *Journal of Business Research*, *57*(5), 478–483. http://doi.org/10.1016/S0148-2963(02)00314-4

Byrne, B. M. (2010). *Structural Equation Modeling with AMOS:Basic Concepts, Applications, and Programming*. *Routledge*. Routledge.

Cai, Z., Huang, Q., Liu, H., & Liang, L. (2016). The moderating role of information technology capability in the relationship between supply chain collaboration and organizational responsiveness Evidence from China. *International Journal of Operations & Production Management*, *36*(10), 1247–1271.

Calantone, R., & Knight, G. (2000). The critical role of product quality in the international performance of industrial firms. *Industrial Marketing Management*, *29*(6), 493–506.

Cao, G., Duan, Y., & Li, G. (2015). Linking Business Analytics to Decision Making Effectiveness: A Path Model Analysis. *IEEE Transactions on Engineering Management*, *62*(3), 384–395. http://doi.org/10.1109/TEM.2015.2441875

Caputo, A., Marzi, G., & Pellegrini, M. M. (2016). The Internet of Things in manufacturing innovation processes. *Business Process Management Journal*, *22*(2), 383–402.

Carter, C. R. (2005). Purchasing social responsibility and firm performance: The key mediating roles of organizational learning and supplier performance. *International Journal of Physical Distribution & Logistics Management*, *35*(3), 177–194.

Casabayó, M., Agell, N., & Sánchez-Hernández, G. (2014). Improved market segmentation by fuzzifying crisp clusters: A case study of the energy market in Spain. *Expert Systems with Applications*, *42*(3), 1637–1643.

Cepeda, G., & Vera, D. (2007). Dynamic capabilities and operational capabilities: A knowledge management perspective. *Journal of Business Research*, *60*(5), 426–437.

Chae, B. (Kevin). (2015). Insights from Hashtag #SupplyChain and Twitter Analytics: Considering Twitter and Twitter Data for Supply Chain Practice and Research. *International Journal of Production Economics*, *165*, 247–259.

Chae, B. (Kevin), & Olson, D. L. (2013). Business analytics for supply chain: A dynamic-capabilities framework. *International Journal of Information Technology & Decision Making*, *12*(01), 9–26. http://doi.org/10.1142/S0219622013500016

Chae, B. (Kevin), Yang, C., & Olson, D. (2014). The impact of supply chain analytics on operational performance: a resource-based view. *International Journal of Production Research*, *52*(16), 4695–4710. http://doi.org/10.1080/00207543.2013.861616

Chae, B. (Kevin), Yang, C., Olson, D., & Sheu, C. (2014). The impact of advanced analytics and data accuracy on operational performance: A contingent resource based theory (RBT) perspective. *Decision Support Systems*, *59*, 119–126.

Chan, H. K., & Chan, F. T. S. (2009). Effect of information sharing in supply chains with flexibility. *International Journal of Production Research*, *47*(1), 213–232.

Chan, H. K., Lacka, E., Yee, R. W. Y., & Lim, M. K. (2015). The role of social media data in operations and production management. *International Journal of Production Research*, *7543*(July 2015), 1–10.

Chan, S. L., Lu, Y., & Wang, Y. (2018). Data-driven cost estimation for additive manufacturing in cybermanufacturing. *Journal of Manufacturing Systems*, *46*, 115–126.

Chapman, S. (2017). The spatial distribution of industries in Great Britain: 2015. Retrieved April 8, 2018, from https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/articles/thespatialdistributionofindustriesingreatbritain/2015

Chen, C. P. L., & Zhang, C. Y. (2014). Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. *Information Sciences*, *275*, 314–347.

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. http://doi.org/10.1080/07421222.2015.1138364.

Chen, H., Chiang, R. H. L., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. *MIS Quarterly: Management Information Systems*, *36*(4), 1165–1188.

Chen, L., & East, F. (2016). The indirect effect in multiple mediators model by structural equation modeling. *European Journal of Business, Economics and Accountancy*, *4*(3), 36–43.

Chen, Y.-S., Lin, Y.-H., Lin, C.-Y., & Chang, C.-W. (2015). Enhancing Green Absorptive Capacity, Green Dynamic Capacities and Green Service Innovation to Improve Firm Performance: An Analysis of Structural Equation Modeling (SEM). *Sustainability (Switzerland)*, *7*(11), 15674–15692. http://doi.org/10.3390/su71115674

Chen, Y., Wang, Y., Nevo, S., Benitez-Amado, J., & Kou, G. (2015). IT capabilities and product innovation performance: The roles of corporate entrepreneurship and competitive intensity. *Information & Management*, *52*(6), 643–657.

Cheng, J.-H., & Lu, K.-L. (2017). Enhancing effects of supply chain resilience: insights from trajectory and resource-based perspectives. *Supply Chain Management*, *22*(4), 329–340.

Chiang, D. M.-H., Lin, C.-P., & Chen, M.-C. (2011). The adaptive approach for storage assignment by mining data of warehouse management system for distribution centres. *Enterprise Information Systems*, *5*(2), 219–234.

Chiasson, P. (2005). Abduction as an aspect of retroduction. *Semiotica*, *153*, 223–242.

Chien, C.-F., Liu, C.-W., & Chuang, S.-C. (2015). Analysing semiconductor manufacturing big data for root cause detection of excursion for yield enhancement. *International Journal of Production Research*, *7543*(March), 1–13.

Chien, C. F., Peng, J. T., & Yu, H. C. (2016). Building energy saving performance indices for cleaner semiconductor manufacturing and an empirical study. *Computers and Industrial Engineering*, *99*, 448–457.

Chircu, A. M., Sultanow, E., & Chircu, F. C. (2014). Cloud Computing for Big Data Entrepreneurship in the Supply Chain: Using SAP HANA for Pharmaceutical Track-and-Trace Analytics. *Services (SERVICES), 2014 IEEE World Congress On*, 450–451.

Cho, J. J.-K., Ozment, J., & Sink, H. (2008). Logistics capability, logistics outsourcing and firm performance in an e-commerce market. *International Journal of Physical Distribution & Logistics Management*, *38*(5), 336–359.

Chowdhury, M., Jayaram, J., & Prajogo, D. (2017). The influence of socialisation and absorptive capacity on buyer’s innovation performance. *International Journal of Production Research*, 1–17. http://doi.org/10.1080/00207543.2017.1346321

Christiansen, J. (2000). *Competitive Innovation Management: Techniques to improve innovation performance*. London, UK: Palgrave macmillan.

Christopher, M. (2016). *Logistics & supply chain management*. Pearson Education Limited.

Chuang, S. H., & Lin, H. N. (2013). The roles of infrastructure capability and customer orientation in enhancing customer-information quality in CRM systems: Empirical evidence from Taiwan. *International Journal of Information Management*, *33*(2), 271–281. http://doi.org/10.1016/j.ijinfomgt.2012.12.003

Chuang, Y. F., Chia, S. H., & Wong, J. Y. (2014). Enhancing order-picking efficiency through data mining and assignment approaches. *WSEAS Transactions on Business and Economics*, *11*(1), 52–64.

Churchil, G. A. (1979). A Paradigm for Developing Better Measures of Marketing Constructs. *Journal of Marketing Research*, *16*(4), 64–73.

Citroen, C. L. (2011). The role of information in strategic decision-making. *International Journal of Information Management, 31*(6), 493–501.

Cobanoglu, C., Warde, B., & Moreo, P. J. (2001). A comparison of mail, fax and web-based survey methods. *International Journal of Market …*, *43*, 405–410. Retrieved from http://www.amstat.org/sections/SRMS/proceedings/papers/2000_065.pdf%5Cnpapers3://publication/uuid/0196826D-91CB-4198-AA8D-0F9CDB74A4BD

Cohen, W., & Levinthal, D. (1990). Absorptive Capacity : A New Perspective on Learning and Innovation. *Administrative Science Quarterly*, *35*(1), 128–152.

Cohen, W. M., & Levinthal, D. A. (1989). Innovation and Learning : The Two Faces of R & D. *The Economic Journal*, *99*(397), 569–596. http://doi.org/10.1111/j.l468-0297.2010.02368.x

Cohen, W. M., & Levinthal, D. A. (1994). Fortune Favors the Prepared Firm. *Management Science*, *40*(2), 227–251. http://doi.org/10.1287/mnsc.40.2.227

Colicchia, C., & Strozzi, F. (2012). Supply chain risk management: a new methodology for a systematic literature review. *Supply Chain Management: An International Journal*, *17*(4), 403–418. http://doi.org/10.1108/13598541211246558

Collins, L. M., & Lanza, S. T. (2010). *Latent class and latent transition analysis: with applications in the social, behavioral, and health sciences*.

Collis, D. J. (1994). Research Note : How Valuable Are Organizational Capabilities ? *Strategic Management Journal*, *15*, 143–152.

Collis, J., & Hussey, R. (2013). *Business research: A practical guide for undergraduate and postgraduate students*. Palgrave macmillan.

Colton, D., & Covert, R. W. (2007). *Designing and Constructing Instrument for Social Research and Evaluation*. John Wiley & Sons, Inc.

Conner, K. R. (1991). A historical comparison of resource-based theory and five schools of thought within industrial organization economics: do we have a new theory of the firm?. *Journal of Management*, *17*(1), 121–154.

Conner, K. R., & Prahalad, C. K. (1996). A Resource-Based Theory of the Firm: Knowledge Versus Opportunism. *Organization Science*, *7*(5), 477–501.

Cooper, D. R., & Schindler, P. S. (2014). *Business Research Methods*. *Business Research Methods*.

Cosic, R., Shanks, G., & Maynard, S. (2015). A business analytics capability framework. *Australasian Journal of Information Systems*, *19*, 5–19.

Costello, A. B., & Osborne, J. W. (2005). Best Practices in Exploratory Factor Analysis : Four Recommendations for Getting the Most From Your Analysis. *Practical Assessment, Research & Education*, *10*, 1–9. http://doi.org/10.1.1.110.9154

Cowles, E. L., & Nelson, E. (2015). *An Introduction to Survey Research and Data Analysis*. Business Expert Press.

Cox, M., & Ellsworth, D. (1997). Application-controlled demand paging for out-of-core visualization. *Proceedings. Visualization '97 (Cat. No. 97CB36155)*, (July), 235–244,. http://doi.org/10.1109/VISUAL.1997.663888

Creswell, J. W. (2014). *Research design : qualitative, quantitative, and mixed methods approaches* (4th ed., Vol. 53). London, UK: Sage Publications, Inc. http://doi.org/10.1017/CBO9781107415324.004

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*(4), 281.

Crossan, M. M., & Apaydin, M. (2010). A multi-dimensional framework of organizational innovation: A systematic review of the literature. *Journal of Management Studies*, *47*(6), 1154–1191. http://doi.org/10.1111/j.1467-6486.2009.00880.x

Cruz-González, J., López-Sáez, P., & Navas-López, J. E. (2015). Absorbing knowledge from supply-chain, industry and science: The distinct moderating role of formal liaison devices on new product development and novelty. *Industrial Marketing Management*, *47*, 75–85. http://doi.org/10.1016/j.indmarman.2015.02.036

Cruz-Jesus, F., Oliveira, T., & Bacao, F. (2012). Digital divide across the European Union. *Information and Management*, *49*(6), 278–291. http://doi.org/10.1016/j.im.2012.09.003

D'Souza, D. E., & Williams, F. P. (2000). Toward a taxonomy of manufacturing flexibility dimensions. *Journal of Operations Management*, *18*(5), 577–593. http://doi.org/10.1016/S0272-6963(00)00036-X

Dale Stoel, M., & Muhanna, W. A. (2009). IT capabilities and firm performance: A contingency analysis of the role of industry and IT capability type. *Information and Management*, *46*(3), 181–189. http://doi.org/10.1016/j.im.2008.10.002

Damanpour, F. (1991). Organizational Innovation: A Meta-Analysis of Effects of Determinants and Moderators. *Academy of Management Journal*, *34*(3), 555–590. http://doi.org/10.2307/256406

Davenport, T. H. (2006). Competing on analytics. Retrieved February 2, 2016, from https://hbr.org/2006/01/competing-on-analytics

Davenport, T. H., & Harris, J. G. (2007). *Competing on analytics: The new science of winning*. *Harvard Business School Press*. Boston, Masssachusetts.

Davenport, T. H., Harris, J. G., De Long, D. W., & Jacobson, A. L. (2001). Data to Knowledge to Results: BUILDING AN ANALYTIC CAPABILITY. *California Management Review*, *43*(2), 117–138. http://doi.org/10.2307/41166078

Davenport, T. H., Leibold, M., & Voelpel, S. (2006). *Strategic management in the innovation economy :Strategy Approaches and Tools for Dynamic Innovation Capabilities*. Wiley.

Davenport, T. H., & Patil, D. J. (2012). Data scientist: the sexiest job of the 21st century: meet the people who can coax treasure out of messy, unstructured data. *Harvard Business Review*, *90*(October), 70–77. http://doi.org/10.1007/978-1-4302-4873-6_9

Davies, D. L., & Bouldin, D. W. (1979). A Cluster Separation Measure. *IEEE Trans. Pattern Anal. Mach. Intell.*, *1*(2), 224–227. http://doi.org/10.1109/TPAMI.1979.4766909

de Beuckelaer, A., & Wagner, S. M. (2012). Small sample surveys: Increasing rigor in supply chain management research. *International Journal of Physical Distribution & Logistics Management*, *42*(7), 615–639. http://doi.org/10.1108/09600031211258129

De Toni, A., & Meneghetti, A. (2000). Traditional and innovative paths towards time-based competition. *International Journal of Production Economics*, *66*(3), 255–268.

De Vaus, D. A. (2002). *Surveys in social research* (5th ed.). Retrieved from internal-pdf://116.65.119.193/De Vaus 2002 Surveys in Social Research.pdf

de Winter, J. C. F., & Dodou, D. (2012). Factor recovery by principal axis factoring and maximum likelihood factor analysis as a function of factor pattern and sample size. *Journal of Applied Statistics, 39*(4), 695–710.

Debortoli, S., Müller, O., & Vom Brocke, J. (2014). Comparing business intelligence and big data skills: A text mining study using job advertisements. *Business and Information Systems Engineering*, *6*(5), 289–300. http://doi.org/10.1007/s12599-014-0344-2

DeGroote, S. E., & Marx, T. G. (2013). The impact of IT on supply chain agility and firm performance: An empirical investigation. *International Journal of Information Management*, *33*(6), 909–916. http://doi.org/10.1016/j.ijinfomgt.2013.09.001

Delen, D., Erraguntla, M., Mayer, R. J., & Wu, C. N. (2011). Better management of blood supply-chain with GIS-based analytics. *Annals of Operations Research*, *185*(1), 181–193. http://doi.org/10.1007/s10479-009-0616-2

Deng, S., & Yeh, T. H. (2011). Using least squares support vector machines for the airframe structures manufacturing cost estimation. *International Journal of Production Economics*, *131*(2), 701–708. http://doi.org/10.1016/j.ijpe.2011.02.019

Department for Business, E. & I. S. (2017). Business population estimates 2017. Retrieved from https://www.gov.uk/government/statistics/business-population-estimates-2017

DeVellis, R. F. (1991). *Scale development : theory and applications*. London: Sage Publications.

Diamantopoulos, A., & Siguaw, J. A. (2006). Formative versus reflective indicators in organizational measure development: A comparison and empirical illustration. *British Journal of Management*, *17*(4), 263–282. http://doi.org/10.1111/j.1467-8551.2006.00500.x

Dillman, D. A., Smyth, J. D., & Christian, L. M. (2014). *Internet, phone, mail, and mixed-mode surveys: the tailored design method*. *John Wiley & Sons, Inc*. John Wiley & Sons, Inc. http://doi.org/10.1017/CBO9781107415324.004

DiMaggio, P. J., & Powell, W. W. (1983). The Iron Cage Revisited : Institutional Isomorphism and Collective Rationality in Organizational. *American Sociological Review*, *48*(2), 147–160.

Dimitrov, D. M. (2010). Testing for factorial invariance in the context of construct Validation. *Measurement and Evaluation in Counseling and Development*, *43*(2), 121–149. http://doi.org/10.1177/0748175610373459

Dobrzykowski, D. D., Leuschner, R., Hong, P. C., & Roh, J. J. (2015). Examining Absorptive Capacity in Supply Chains: Linking Responsive Strategy and Firm Performance. *Journal of Supply Chain Management*, *51*(4), 3–28.

Dolničar, S. (2003). Using cluster analysis for market segmentation - typical misconceptions, established methodological weaknesses and some recommendations for improvement. *Australasian Journal of Market Research*, *11*(2), 5–12.

Donovan, P. O., Leahy, K., Bruton, K., & Sullivan, D. T. J. O. (2015). Big data in manufacturing : a systematic mapping study. *Journal of Big Data*. http://doi.org/10.1186/s40537-015-0028-x

Droge, C., Jayaram, J., & Vickery, S. K. (2004). The effects of internal versus external integration practices on time-based performance and overall firm performance. *Journal of Operations Management*, *22*(6), 557–573.

Duan, Y., & Cao, G. (2015). Understanding the Impact of Business Analytics on Innovation. *ECIS 2015 Completed Research Papers. Paper 40*. Retrieved from http://aisel.aisnet.org/ecis2015_cr/40

Dubey, R., Gunasekaran, A., Childe, S. J., Wamba, S. F., & Papadopoulos, T. (2015). The impact of big data on world-class sustainable manufacturing. *International Journal of Advanced Manufacturing Technology*, *84*, 631–645. http://doi.org/10.1007/s00170-015-7674-1

Dunn, J. C. (1973). A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters. *Journal of Cybernetics*, *3*(3), 32–57.

Durach, C. F., Kembro, J., & Wieland, A. (2017). A new paradigm for systematic literature reviews in supply chain management. *Journal of Supply Chain Management*, *53*(October), 67–85. http://doi.org/10.1111/jscm.12145

Dutta, D., & Bose, I. (2015). Managing a big data project: The case of Ramco Cements limited. *International Journal of Production Economics*, *165*, 293–306.

Dy, J. G., & Brodley, C. E. (2004). Feature Selection for Unsupervised Learning. *The Journal of Machine Learning Research*, *5*(4), 845–889. Retrieved from http://portal.acm.org/citation.cfm?id=1016787&dl=GUIDE,

Dyer, J. H., & Singh, H. (1998). The Relational View : Cooperative Strategy and Sources of Interorganizational Competitive Advantage. *The Academy of Management Review*, *23*(4), 660–679.

Easton, G. (2010). Critical realism in case study research. *Industrial Marketing Management*, *39*(1), 118–128. http://doi.org/10.1016/j.indmarman.2008.06.004

Edwards, P. K., O'Mahoney, J., & Vincent, S. (2014). *Studying Organisations using critical realism : A Practical Guide*. *Oxford University Press* (1st ed.). Oxford, UK.

Eisenhardt, K. M., & Martin, A. J. (2000). Dynamic capabilities: what are they? *Strategic Management Journal*, *21*, 1105–1121. http://doi.org/10.1002/1097-0266(200010/11)21:10/11<1105::AID-SMJ133>3.0.CO;2-E

Enders, C. K. (2010). Applied missing data analysis. *Methodology in the Social Sciences*. http://doi.org/10.1017/CBO9781107415324.004

Erevelles, S., Fukawa, N., & Swayne, L. (2015). Big Data consumer analytics and the transformation of marketing. *Journal of Business Research*. http://doi.org/10.1016/j.jbusres.2015.07.001

Ettlie, J. E., & Pavlou, P. A. (2006). Technology-based new product development partnerships. *Decision Sciences*, *37*(2), 117–147. http://doi.org/10.1111/j.1540-5915.2006.00119.x

Eusébio, C., João Carneiro, M., Kastenholz, E., & Alvelos, H. (2017). Social tourism programmes for the senior market: a benefit segmentation analysis. *Journal of Tourism and Cultural Change*, *15*(1), 59–79. http://doi.org/10.1080/14766825.2015.1117093

Evely, A. C., Fazey, I., Pinard, M., & Lambin, X. (2008). The influence of philosophical perspectives in integrative research: A conservation case study in the Cairngorms National Park. *Ecology and Society*, *13*(2), 52. http://doi.org/Artn 52

Fabrigar, L. L. R., Wegener, D. D. T., MacCallum, R. R. C., & Strahan, E. E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, *4*(3), 272–299. http://doi.org/10.1037//1082-989X.4.3.272

Fahimnia, B., Sarkis, J., & Davarzani, H. (2015). Green supply chain management: A review and bibliometric analysis. *International Journal of Production Economics*, *162*, 101–114.

Fawcett, S. E. ., & Waller, M. A. (2014). Supply Chain Game Changers—Mega, Nano, and Virtual Trends—And Forces That Impede Supply Chain Design (i.e., Building a Winning Team). *Journal of Business Logistics*, *35*(3), 157–164.

Fawcett, S. E., Wallin, C., Allred, C., Fawcett, A. M., & Magnan, G. M. (2011). Information technology as an enabler of Supply Chain Collaboration:A dynamic-capabilities perspectives. *Journal of Supply Chain Management*, *47*(1), 38–59.

Fayard, D., Lee, L. S., Leitch, R. A., & Kettinger, W. J. (2012). Effect of internal cost management, information systems integration, and absorptive capacity on interorganizational cost management in supply chains. *Accounting, Organizations and Society*, *37*(3), 168–187. http://doi.org/10.1016/j.aos.2012.02.001

Ferdows, K., & De Meyer, A. (1990). Lasting improvements in manufacturing performance: In search of a new theory. *Journal of Operations Management*, *9*(2), 168–184.

Few, S. (2009). *Now You See It: Simple Visualization Techniques for Quantitative Analysis*. *Distribution*. Analytics Press. http://doi.org/10.1016/S0262-4079(06)60844-0

Field, A. (2013). *Discovering statistics using IBM SPSS statistics* (4th ed.). London, UK: Sage Publications.

Fildes, R., Goodwin, P., Lawrence, M., & Nikolopoulos, K. (2009). Effective forecasting and judgmental adjustments: an empirical evaluation and strategies for improvement in supply-chain planning. *International Journal of Forecasting*, *25*(1), 3–23.

Fiol, C. M. (1991). Managing Culture as a Competitive Resource: An Identity-Based View of Sustainable Competitive Advantage. *Journal of Management*, *17*(1), 191–211.

Fiol, C. M. (2001). Revisting an identity-based view of sustainable competitive advantage. *Journal of Management*, *27*, 691–699. http://doi.org/10.1177/014920630102700606

Fleetwood, S. (2004). *The ontology of organisation and management studies*. *Realist Applications in Organisation and Management Studies*. London, UK: Routledge.

Floyd, F. J., & Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment*, *7*(3), 286–299.

Flynn, B. (1990). Empirical research methods in operations management. *Journal of Operations Management*, *9*(2), 250–284. http://doi.org/10.1016/0272-6963(90)90098-X

Ford, K. J., MacCallum, R. C., & Tait, M. (1986). The Application of Exploratory Factor Analysis in Applied Psychology: a Critical Review and Analysis. *Personnel Psychology*, *39*(2), 291–314. http://doi.org/10.1111/j.1744-6570.1986.tb00583.x

Formative indicators: Introduction to the special issue. (2008). *Journal of Business Research*. http://doi.org/10.1016/j.jbusres.2008.01.008

Fornell, C., Johnson, M. D., Anderson, E. W., & Bryant, B. E. (1996). The American Customer Satisfaction Index : Nature, Purpose, and Findings, *60*(4), 7–18.

Forza, C. (2002). Survey research in operations management: a process-based perspective. *International Journal of Operations & Production Management*, *22*(2), 152–194.

Forza, C., & Filippini, R. (1998). TQM impact on quality conformance and customer satisfaction: A causal model. *International Journal of Production Economics*, *55*(1), 1–20.

Fosso Wamba, S., & Akter, S. (2015). *Big data analytics for supply chain management: A literature review and research agenda*. (J. Barjis, R. Pergl, & E. Babkin, Eds.), *Lecture Notes in Business Information Processing* (Vol. 231). Cham: Springer International Publishing.

Fowler Jr, F. J. (2014). *Survey research methods.* (5th ed.). SAGE Publications Ltd.

Francas, D., Löhndorf, N., & Minner, S. (2011). Machine and labor flexibility in manufacturing networks. *International Journal of Production Economics*, *131*(1), 165–174.

Frankel, R., Naslund, D., & Bolumole, Y. (2005). The “White Space” of Logistics Research:

A Look at the Role of Methods Usage. *Journal of Business Logistics*, *26*(2), 185–208.

Fredericks, E. (2005). Infusing flexibility into business-to-business firms: A contingency theory and resource-based view perspective and practical implications. *Industrial Marketing Management*, *34*(6 SPEC. ISS.), 555–565.

Fredriksson, P., & Gadde, L. E. (2005). Flexibility and rigidity in customization and build-to-order production. *Industrial Marketing Management*, *34*(7 SPEC. ISS.), 695–705. http://doi.org/10.1016/j.indmarman.2005.05.010

Fritz, M. S., & Mackinnon, D. P. (2007). Required Sample Size to Detect the Mediated Effect. *Psychological Science*, *18*(3), 233–239. http://doi.org/10.1111/j.1467-9280.2007.01882.x

Frizzo-Barker, J., Chow-White, P. A., Mozafari, M., & Ha, D. (2016). An empirical study of the rise of big data in business scholarship. *International Journal of Information Management*, *36*(3), 403–413.

Fynes, B., & De Búrca, S. (2005). The effects of design quality on quality performance. *International Journal of Production Economics*, *96*(1), 1–14.

Gamble, J., Jr, A. T., & Peteraf, M. (2014). *Essentials of Strategic Management: The Quest for Competitive Advantage* (4th ed.). McGraw-Hill Education.

Gammelgaard, J., McDonald, F., Tüselmann, H., Dörrenbächer, C., & Stephan, A. (2011). Effective autonomy, organisational relationships and skilled jobs in subsidiaries. *Management Research Review*, *34*(4), 366–385.

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, *35*(2), 137–144. http://doi.org/10.1016/j.ijinfomgt.2014.10.007

Gao, D., Xu, Z., Ruan, Y. Z., & Lu, H. (2016). From a systematic literature review to integrated definition for sustainable supply chain innovation (SSCI). *Journal of Cleaner Production*. http://doi.org/10.1016/j.jclepro.2016.11.153

Gartner. (2014). Gartner Predicts Business Intelligence and Analytics Will Remain Top Focus for CIOs Through 2017. Retrieved August 18, 2016, from http://www.gartner.com/newsroom/id/2637615

Garvin, D. A. (1984). Product quality: An important strategic weapon. *Business Horizons*, *27*(3), 40–43.

Gaskin, J., & Lim, J. (2017). CFA Tool. AMOS Plugin. Retrieved from http://statwiki.kolobkreations.com/

Ge, X., & Jackson, J. (2014). The Big Data Application Strategy for Cost Reduction in Automotive Industry. *SAE International Journal of Commercial Vehicles*, *7*(2), 588–598.

Genpact. (2015). *Driving supply chain excellence through Data-to-Action Analytics*. Retrieved from http://www.genpact.com/downloadable-content/insight/driving-supply-chain-excellence-through-data-to-action-analytics.pdf

George, G., Zahra, S. A., Wheatley, K. K., & Khan, R. (2001). The effects of alliance portfolio characteristics and absorptive capacity on performance A study of biotechnology firms. *Journal of High Technology Management Research*, *12*(2), 205–226.

Gerwin, D. (1993). Manufacturing Flexibility: A Strategic Perspective. *Management Science*, *39*(4), 395–410.

Gill, J., & Johnson, P. (2002). *Research Methods for Managers* (3rd ed.). London, UK: SAGE Publications Ltd.

Gioia, D. A., & Pitre, E. (1990). Multiparadigm perspectives on theory building. *Academy of Managment Review*, *15*(4), 584–602.

Gjerde, K. A. P., & Slotnick, S. A. (2004). Quality and reputation: The effects of external and internal factors over time. *International Journal of Production Economics*, *89*(1), 1–20.

Glenn Richey Jr, R., Chen, H., Upreti, R., Fawcett, S. E., & Adams, F. G. (2009). The moderating role of barriers on the relationship between drivers to supply chain integration and firm performance. *International Journal of Physical Distribution & Logistics Management*, *39*(10), 826–840.

Gligor, D. M., & Holcomb, M. C. (2012). Antecedents and consequences of supply chain agility: Establishing the link to firm performance. *Journal of Business Logistics*, *33*(4), 295–308.

Gnambs, T., & Kaspar, K. (2015). Disclosure of sensitive behaviors across self-administered survey modes: a meta-analysis. *Behavior Research Methods*, *47*(4), 1237–1259.

Golden, W., & Powell, P. (1999). Exploring inter-organisational systems and flexibility in Ireland: a case of two value chains. *International Journal of Agile Management Systems*, *1*(3), 169–176. http://doi.org/10.1108/14654659910296544

Goles, T., & Hirschheim, R. (2000). The paradigm is dead, the paradigm is dead… long live the paradigm: the legacy of Burrell and Morgan. *Omega*, *28*, 249–268.

Golicic, S. L., & Smith, C. D. (2013). a Meta-Analysis of Environmentally Sustainable Supply Chain Management Practices and Firm Performance. *Journal of Supply Chain Management*, *49*(2), 78–95.

Gopalakrishnan, S., & Damanpour, F. (1997). A review of innovation research in economics, sociology and technology management. *International Journal Management Science*, *25*(1), 15–28.

Gorla, N., Somers, T. M., & Wong, B. (2010). Organizational impact of system quality, information quality, and service quality. *Journal of Strategic Information Systems*, *19*(3), 207–228.

Gosling, J., Jia, F., Gong, Y., & Brown, S. (2017). The role of supply chain leadership in the learning of sustainable practice: Toward an integrated framework. *Journal of Cleaner Production*, *140*, 239–250. http://doi.org/10.1016/j.jclepro.2016.09.101

Gosling, S. D., Vazire, S., Srivastava, S., & John, O. P. (2004). Should We Trust Web-Based Studies? A Comparative Analysis of Six Preconceptions About Internet Questionnaires. *American Psychologist*, *59*(2), 93–104. http://doi.org/10.1037/0003-066X.59.2.93

Gower, J. C. (1971). A General Coefficient of Similarity and Some of Its Properties. *Biometrics*, *27*(4), 857–871.

Grant, R. M. (1996a). Prospering in Dynamically-Competitive Environments: Organizational Capability as Knowledge Integration. *Organization Science*, *7*(4), 375–387.

Grant, R. M. (1996b). Towards a knowledge-based theory of the firm. *Strategic Management Journal*, *17*, *Winter*, 109–122. http://doi.org/10.2307/2486994

Grekova, K., Calantone, R. J., Bremmers, H. J., Trienekens, J. H., & Omta, S. W. F. (2015).

How environmental collaboration with suppliers and customers influences firm performance: evidence from Dutch food and beverage processors. *Journal of Cleaner Production*, *112*, 1861–1871. http://doi.org/10.1016/j.jclepro.2015.03.022

Grewal, R., & Slotegraaf, R. J. (2007). Embeddedness of organizational capabilities. *Decision Sciences*, *38*(3), 451–488. http://doi.org/10.1111/j.1540-5915.2007.00166.x

Guba, E. G., & Lincoln, Y. S. (1994). Competing Paradigms in Qualitative Research. *Handbook of Qualitative Research*, 105–117. http://doi.org/http://www.uncg.edu/hdf/facultystaff/Tudge/Guba%20&%20Lincoln%201994.pdf

Gudfinnsson, K., Strand, M., & Berndtsson, M. (2015). Analyzing business intelligence maturity. *Journal of Decision Systems*, *24*(1), 37–54.

Gunasekaran, A., Papadopoulos, T., Dubey, R., Wamba, S. F., Childe, S. J., Hazen, B., & Akter, S. (2017). Big data and predictive analytics for supply chain and organizational performance. *Journal of Business Research*, *70*, 308–317.

Gunasekaran, A., Patel, C., & Tirtiroglu, E. (2001). *Performance measures and metrics in a supply chain environment*. *International Journal of Operations & Production Management (ABS2015:4)* (Vol. 21). http://doi.org/10.1108/01443570110358468

Gunasekaran, A., Yusuf, Y. Y., Adeleye, E. O., & Papadopoulos, T. (2018). Agile manufacturing practices: the role of big data and business analytics with multiple case studies. *International Journal of Production Research*, *56*(1–2), 385–397.

Gunasekaran, & Kobu. (2007). Performance measures and metrics in logistics and supply chain management: a review of recent literature (1995–2004) for research and applications. *International Journal of Production Research*, *45*(12), 2819–2840.

Gunday, G., Ulusoy, G., Kilic, K., & Alpkan, L. (2011). Effects of innovation types on firm performance. *International Journal of Production Economics*, *133*(2), 662–676.

Gupta, M., & George, J. F. (2016). Toward the development of a big data analytics capability. *Information & Management*, *53*(8), 1049–1064.

Gupta, Y. P., & Goyal, S. (1989). Flexibility of manufacturing systems: Concepts and measurements. *European Journal of Operational Research*, *43*(2), 119–135.

Gupta, Y. P., & Somers, T. M. (1992). Measurement of Manufacturing Flexibility. *European Journal of Operational Research*, *60*(2), 166–182.

Gustavsson, M., & Wänström, C. (2009). Assessing information quality in manufacturing planning and control processes. *International Journal of Quality & Reliability Management*, *26*(4), 325–340. http://doi.org/10.1108/02656710910950333

Hagenaars, J. A., & McCutcheon, A. L. (2002). *Applied Latent Class Analysis*. Cambridge University Press. http://doi.org/10.1017/CBO9781107415324.004

Hagtvet, K. A., & Nasser, F. M. (2004). How Well Do Item Parcels Represent Conceptually Defined Latent Constructs? A Two-Facet Approach. *Structural Equation Modeling*, *11*(2), 168–193. http://doi.org/10.1207/s15328007sem1102

Hague, R., Reeves, P., & Jones, S. (2016). Mapping UK Research and Innovation in Additive manufacturing. Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/505246/CO307_Mapping_UK_Accessible.pdf

Hahn, G. J., & Packowski, J. (2015). A Perspective on Applications of In-memory Analytics in Supply Chain Management. *Decision Support Systems*, *76*, 45–52.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (1998). *Multivariate Data Analysis*. *Prentice-Hall, Inc* (Vol. 1). http://doi.org/10.1038/259433b0

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate Data Analysis* (7th ed.). Pearson Prentice Hall. http://doi.org/10.1016/j.ijpharm.2011.02.019

Hair, J. F. J., Hult, G. T. M., Ringle, C., & Sarstedt, M. (2014). *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)*. *SAGE Publications, Inc.*

Hallavo, V. (2015). Superior performance through supply chain fit: a synthesis. *Supply Chain Management: An International Journal*, *20*(1), 71–82.

Halper, F., & Krishnan, K. (2014). TDWi Big Data Maturity Model Guide. *TDWI Resarch*, *2013–2014*, 1–20. Retrieved from http://www.pentaho.com/sites/default/files/uploads/resources/tdwi_big_data_maturity_model_guide_2013.pdf

Halter, O., & Rao, A. (2013). Capitalizing on the promise of Big Data. Retrieved August 22, 2016, from https://www.pwc.com/us/en/increasing-it-effectiveness/assets/capitalizing-on-the-promise-of-big-data.pdf

Hannula, M., & Pirttimäki, V. (2003). Business intelligence empirical study on the top 50 Finnish companies. *Journal of American Academy of Business*, *2*, 593–599. Retrieved from http://www.umsl.edu/~lacitym/bi1.pdf

Hardie, M., & Banks, A. (2014). The Changing Shape of UK Manufacturing. *Office for National Statistics*, (October), 1–20. Retrieved from http://www.ons.gov.uk/ons/dcp171766_381512.pdf

Harris, D. (2014). Google Has Open-Sourced a Tool for Inferring Cause From Correlations. Retrieved from https://gigaom.com/2014/09/11/google-has-open-sourced-a-tool-for-inferring-cause-from-correlations/

Haryadi, A. F., Hulstijn, J., Wahyudi, A., van der Voort, H., & Janssen, M. (2016). Antecedents of big data quality: An empirical examination in financial service organizations. *IEEE International Conference on Big Data*, 116–121.

Haughton, D., Legrand, P., & Woolford, S. (2009). Review of three latent class cluster analysis packages: Latent gold, poLCA, and MCLUST. *American Statistician*, *63*(1), 81–91.

Hayes, A. F. (2013). *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach*. London, UK: The Guilford Press.

Hazen, B. T., Boone, C. A., Ezell, J. D., & Jones-Farmer, L. A. (2014). Data quality for data science, predictive analytics, and big data in supply chain management: An introduction to the problem and suggestions for research and applications. *International Journal of Production Economics*, *154*, 72–80.

Hazen, B. T., Skipper, J. B., Ezell, J. D., & Boone, C. A. (2016). Big Data and Predictive Analytics for Supply Chain Sustainability: A Theory-driven Research Agenda. *Computers & Industrial Engineering*. http://doi.org/10.1016/j.cie.2016.06.030

Hedrick, T. E., Bickman, L., & Rog, D. J. (1993). *Applied research design: A practical guide* (Vol. 32). Sage Publications.

Heizer, J., Render, B., & Munson, C. (2017). *Operations management: sustainability and supply chain management* (12th ed.). London, UK: Pearson Education Limited.

Hennig, C., Meila, M., Murtagh, F., & Rocci, R. (2015). *Handbook of Cluster Analysis*. Taylor & Francis Group, LLC.

Hill, C. W. L., Jones, G. R., & Schilling, M. A. (2015a). *Strategic management: An integrated approach* (11th ed.). Cengage Learning. Retrieved from http://books.google.com.co/books?id=0588ekqiqQAC&printsec=frontcover&hl=es&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false

Hill, C. W. L., Jones, G. R., & Schilling, M. A. (2015b). *Strategic management: An integrated approach* (11th ed.).

Hinkin, T. . R., Tracey, J. . B., & Enz, C. A. (1997). Scale construction: Developing reliable and valid measurement instruments. *Journal of Hospitality and Tourism Research*, *21*(1), 100–120.

Hofmann, E. (2015). Big data and supply chain decisions: the impact of volume, variety and velocity properties on the bullwhip effect. *International Journal of Production Research*, *7543*(December 2015), 1–19.

Hollos, D., Blome, C., & Foerstl, K. (2012). Does sustainable supplier co-operation affect performance? Examining implications for the triple bottom line. *International Journal of Production Research*, *50*(11), 2968–2986.

Hosseini, S., & Khaled, A. Al. (2016). A hybrid ensemble and AHP approach for resilient supplier selection. *Journal of Intelligent Manufacturing*, 1–22.

Howson, C. (2015). *IT Score Overview for BI and Analytics*. Retrieved from http://www.gartner.com/document/3136418#

Hoyle, R. H. (2012). *Handbook of structural equation modeling*. New York, USA: The Guilford Press.

Hsu, C.-C., Kannan, V. R., Tan, K.-C., & Leong, G. K. (2008). Information sharing, buyer-supplier relationships, and firm performance: A multi-region analysis. *International Journal of Physical Distribution & Logistics Management*, *38*(4), 296–310.

Hu, H., Wen, Y., Chua, T. S., & Li, X. (2014). Toward Scalable Systems for Big Data Analytics: A Technology Tutorial. *Access, IEEE*, *2*, 652–687.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55.

Hubley, J. (2001). Data Quality: The Foundation for Business Intelligence.

Huo, B., Qi, Y., Wang, Z., & Zhao, X. (2014). The impact of supply chain integration on firm performance : The moderating role of competitive strategy. *Supply Chain Management: An International Journal*, *19*(4), 369–384.

Hurley, R. F., & Hult, G. T. M. (1998). Innovation, Market Orientation, and Organizational Learning: An Integration and Empirical Examination. *Journal of Marketing*, *62*(3), 42.

Hurwitz, J., Nugent, A., Halper, F., & Kaufman, M. (2013). *Big Data For Dummies*. John Wiley & Sons, Hoboken, New Jersey.

Hutcheson, G. D., & Sofroniou, N. (1999). *The Multivariate Social Scientist: an introduction to generalized linear models*. SAGE Publications Ltd.

Iacobucci, D. (2009). Everything you always wanted to know about SEM (structural equations modeling) but were afraid to ask. *Journal of Consumer Psychology*, *19*(4), 673–680.

IBM. (2013). IBM Narrows Big Data Skills Gap By Partnering With More Than 1,000 Global Universities. *Education Letter*, 1–3. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=plh&AN=201308140900PR.NEWS.USPR.NY63777&site=eds-live

IDC. (2013). CSC Big data maturity tool. Retrieved August 5, 2016, from http://csc.bigdatamaturity.com/

Inman, R. R., Blumenfeld, D. E., Huang, N., & Li, J. (2003). Designing production systems for quality: Research opportunities from an automotive industry perspective. *International Journal of Production Research*, *41*(9), 1953–1971.

Ipcc. (2013). Big Data Fueled Process Management of Supply Risk: Sending, Prediction, Evaluation And Mitigation. In *Climate Change 2013: The Physical Science Basis* (Vol. 53, pp. 1–36).

Jabar, J., Soosay, C., & Santa, R. (2011). Organisational learning as an antecedent of technology transfer and new product development: A study of manufacturing firms in Malaysia. *Journal of Manufacturing Technology Management*, *22*(1), 25–45.

Jain, A., Jain, P. K., Chan, F. T. S., & Singh, S. (2013). A review on manufacturing flexibility. *International Journal of Production Research*, *51*(19), 5946–5970.

Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, *31*(8), 651–666.

Jain, S., Shao, G., & Shin, S. J. (2017). Manufacturing data analytics using a virtual factory representation. *International Journal of Production Research*, *55*(18), 5450–5464.

Jain, S., Triantis, K. P., & Liu, S. (2011). Manufacturing performance measurement and target setting: A data envelopment analysis approach. *European Journal of Operational Research*, *214*(3), 616–626.

Jajuga, K., Bak, A., & Walesiak, M. (2000). On the general distance measure, (1), 104–109.

Jamehshooran, B. G., Shaharoun, A. M., & Norehan, H. H. (2015a). Assessing Supply Chain Performance through Applying the SCOR Model. *International Journal of Supply Chain Management*, *4*(1), 1–11.

Jamehshooran, B. G., Shaharoun, A. M., & Norehan, H. H. (2015b). The moderating effect of web service on the relationship between business analytics and supply chain performance. *Journal of Theoretical and Applied Information Technology*, *76*(1), 97–108.

Jansen, J. J. P., Van Den Bosch, F. A. J., & Volberda, H. W. (2006). Exploratory Innovation, Exploitative Innovation, and Performance: Effects of Organizational Antecedents and Environmental Moderators. *Management Science*, *52*(11), 1661–1674.

Janssen, M., Estevez, E., & Janowski, T. (2014). Interoperability in big, open, and linked data-organizational maturity, capabilities, and data portfolios. *Computer*, *47*(10), 44–49.

Janssen, M., van der Voort, H., & Wahyudi, A. (2017). Factors influencing big data decision-making quality. *Journal of Business Research*, *70*, 338–345.

Jean, R.-J. `Bryan', Sinkovics, R. R., & Kim, D. (2008). Information technology and organizational performance within international business to business relationships A review and an integrated conceptual framework. *International Marketing Review*, *25*(5), 563–583.

Ji, W., & Wang, L. (2017). Big data analytics based fault prediction for shop floor scheduling. *Journal of Manufacturing Systems*, *43*, 187–194.

Joshi, A. W., & Sharma, S. (2004). Customer Knowledge Development: Antecedents and Impact on New Product Performance. *Journal of Marketing*, *68*(4), 47–59.

Kache, F. (2015). *Dealing with digital information richness in supply chain management – A review and a Big Data Analytics approach Supply Chain Management*. *kassel university press*. The University of Kassel. Retrieved from http://www.uni-kassel.de/upress/online/OpenAccess/978-3-86219-926-6.OpenAccess.pdf

Kache, F., & Seuring, S. (2017). Challenges and opportunities of digital information at the intersection of Big Data Analytics and supply chain management. *International Journal of Operations & Production Management*, *37*(1), 10–36.

Kaiser, H. F., & Rice, J. (1974). Little Jiffy, Mark IV. *Educational and Psychological Measurement*, *34*(1), 111–117.

Kambatla, K., Kollias, G., Kumar, V., & Grama, A. (2014). Trends in big data analytics. *Journal of Parallel and Distributed Computing*, *74*(7), 2561–2573.

Kang, Y., Park, I., Rhee, J., & Lee, Y. (2016). MongoDB-based Repository Design for IoT-generated RFID / Sensor Big Data. *IEEE Sensors Journal*, *16*(2), 485–497.

Kaufman, L., & Rousseeuw, P. J. (1990). *Finding Groups in Data: An Introduction to Cluster Analysis*. http://doi.org/10.1007/s13398-014-0173-7.2

Kauppi, K., Brandon-Jones, A., Ronchi, S., & van Raaij, E. M. (2013). Tools without skills:Exploring the moderating effect of absorptive capacity on the relationship between e-purchasing tools and category performance. *International Journal of Operations & Production Management*, *33*(7), 828–857. http://doi.org/10.2966/scrip.

Kawalek, J. P. (2008). *Rethinking Information Systems in Organizations: integrating organizational problem solving*. Routledge.

Kaydos, W. J. (2000). *Operational Performance Measurement : Increasing Total Productivity* (Vol. 8). London: Taylor & Francis Group, LLC.

Kaynak, H. (2003). The relationship between total quality management practices and their effects on firm performance. *Journal of Operations Management*, *21*(4), 405–435.

Kenett, R. S., & Shmueli, G. (2017). *Information Quality: The Potential of Data and Analytics to Generate Knowledge*. John Wiley & Sons.

Kerlinger, F. N. (1986). *Foundations of behavioral research* (3rd ed.). Holt, Rinehart, and Winston. Klein.

Khan K, A., Bakkappa, B., Metri, B. A., & Sahay, B. S. (2009). Impact of agile supply chains' delivery practices on firms' performance: cluster analysis and validation. *Supply Chain Management: An International Journal*, *14*(1), 41–48.

Khan, Z., & Nicholson, J. D. (2014). An investigation of the cross-border supplier development process: Problems and implications in an emerging economy. *International Business Review*, *23*(6), 1212–1222.

Kim, D. Y., Kumar, V., & Kumar, U. (2012). Relationship between quality management practices and innovation. *Journal of Operations Management*, *30*(4), 295–315.

Kim, H., Hur, D., & Schoenherr, T. (2015). When Buyer-Driven Knowledge Transfer Activities Really Work: A Motivation-Opportunity-Ability Perspective. *Journal of Supply Chain Management*, *51*(3), 33–60.

Kim, S. H., & Lee, J. (2015). Absorptive capacities and supply chain performance: Focus on Korean firms and suppliers. *Information (Japan)*, *18*(3), 825–830.

Kim, S. W. (2006). Effects of supply chain management practices, integration and competition capability on performance. *Supply Chain Management: An International Journal*, *11*(3), 241–248.

Kim, S. W. (2009). An investigation on the direct and indirect effect of supply chain integration on firm performance. *International Journal of Production Economics*, *119*(2), 328–346.

Kirchoff, J. F., Tate, W. L., & Mollenkopf, D. A. (2016). The impact of strategic organizational orientations on green supply chain management and firm performance. *International Journal of Physical Distribution & Logistics Management*, *46*(3), 269–292.

Kirk, A. (2012). *Data Visualization: a successful design process*. Packt Publishing Ltd.

Kiron, D., & Shockley, R. (2011). Creating business value with analytics. *MIT Sloan Management Review*, *53*(1), 57–63. Retrieved from http://sloanreview.mit.edu/article/creating-business-value-with-analytics/

Klassen, R. D., & Jacobs, J. (2001). Experimental comparison of Web, electronic and mail survey technologies in operations management. *Journal of Operations Management*, *19*(6), 713–728.

Klassen, R. D., & Vereecke, A. (2012). Social issues in supply chains: Capabilities link responsibility, risk (opportunity), and performance. *International Journal of Production Economics*, *140*(1), 103–115.

Kline, P. (1994). *An easy guide to factor analysis*. *Routledge*. http://doi.org/10.1016/0191-8869(94)90040-X

Kline, R. B. (2015). *Principles and practices of structural equation modelling*. *The Guilford Press* (4th ed.). London.

Knoppen, D., Johnston, D., & Sáenz, M. J. (2015). Supply chain relationships as a context for learning leading to innovation. *The International Journal of Logistics Management*, *26*(3), 543–567.

Kor, Y. Y., & Mahoney, J. T. (2000). Penrose's resource-based approach: the process and product of research creativity*. *Journal of Management Studies*, *37*(1), 109–139.

Kor, Y. Y., & Mahoney, J. T. (2004). Edith Penrose' s (1959) Contributions to the Resource-based View of Strategic Management. *Journal of Management Studies*, *41*(1), 183–191.

Koste, L., & Malhotra, M. (1999). A theoretical framework for analyzing the dimensions of manufacturing flexibility. *Journal of Operations Management*, *18*, 75–93.

Kostopoulos, K., Papalexandris, A., Papachroni, M., & Ioannou, G. (2011). Absorptive capacity, innovation, and financial performance. *Journal of Business Research*, *64*(12), 1335–1343.

Koufterosa, X. A., Vonderembse, M. A., & Doll, W. J. (1998). Developing measures of time based manufacturing. *Journal of Operations Management*, *16*(1), 21–41.

Kraaijenbrink, J., Spender, J. C., & Groen, A. J. (2010). The Resource-based view: A review and assessment of its critiques. *Journal of Management*, *36*(1), 349–372.

Kraus, C., & Valverde, R. (2014). A data warehouse design for the detection of fraud in the supply chain by using the benford's law. *American Journal of Applied Sciences*, *11*(9), 1507–1518.

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

Kroes, J. R., & Ghosh, S. (2010). Outsourcing congruence with competitive priorities: Impact on supply chain and firm performance. *Journal of Operations Management*, *28*(2), 124–143.

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. http://doi.org/10.1016/j.ijinfomgt.2014.02.002

Laari, S., Töyli, J., Solakivi, T., & Ojala, L. (2015). Firm performance and customer-driven green supply chain management. *Journal of Cleaner Production*, *112*, 1960–1970.

Lane, P. J., Koka, B. R., & Pathak, S. (2006). The reification of Absorptive Capacity: A critical review and rejuvenatino of the construct. *Academy of Management Review*, *31*(4), 833–863.

Lane, P. J., & Lubatkin, M. (1998). Relative absorptive capacity and interorganizational learning. *Strategic Management Journal*, *19*(5), 461–477.

Laney, D. (2015). *Innovating With Analytics: 40 Real World Examples in 40 Minutes*. Retrieved from http://www.gartner.com//it/content/3163300/3163318/innovating_with_analytics_laney.pdf?userId=84223790

Laursen, G. H. N., & Thorlund, J. (2010). *Business Analytics for Managers: Taking Business Intelligence Beyond Reporting*. *Wiley and SAS business series*. New Jersey: John Wiley & Sons, Inc. Retrieved from http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:BUSINESS+ANALYTICS#7

Lavalle, S., Hopkins, M. S., Lesser, E., Shockley, R., & Kruschwitz, N. (2010). Analytics : The New Path to Value. *MIT Sloan Management Review*, 1–24. Retrieved from http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Analytics+:+The+New+Path+to+Value#0

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.

Lawson, B., & Potter, A. (2012). Determinants of knowledge transfer in inter-firm new product development projects. *International Journal of Operations & Production Management*, *32*(10), 1228–1247.

Lawson, B., Tyler, B. B., & Potter, A. (2015). Strategic Suppliers' Technical Contributions to New Product Advantage: Substitution and Configuration Options. *Journal of Product Innovation Management*, *32*(5), 760–776.

Lazer, D., Kennedy, R., King, G., & Vespignani, A. (2014). The Parable of Google Flu: Traps in Big Data Analysis. *Science*, *343*(6167), 1203–1205.

Lee, C.-H., Yang, H.-C., Cheng, S.-C., & Tsai, S.-W. (2015). A Hybrid Big Data Analytics Method for Yield Improvement in Semiconductor Manufacturing. *Proceedings of the ASE BigData & SocialInformatics*, 9.

Lee, E.-S., & Song, D.-W. (2015). The effect of shipping knowledge and absorptive capacity on organizational innovation and logistics value. *The International Journal of Logistics Management*, *26*(2), 218–237.

Lee, H. L. (2018). Big Data and the Innovation Cycle. *Production and Operations Management*, *0*(0), 1–5. http://doi.org/10.1111/poms.12845

Lee, J., Kao, H.-A., & Yang, S. (2014). Service Innovation and Smart Analytics for Industry 4.0 and Big Data Environment. *Procedia CIRP*, *16*, 3–8.

Lee, J., Lapira, E., Bagheri, B., & Kao, H. (2013). Recent advances and trends in predictive manufacturing systems in big data environment. *Manufacturing Letters*, *1*(1), 38–41.

Lee, S. (2003). *Data Quality in Organizational Context : a Case Study*.

Lee, Y. W., Pipino, L. L., Funk, J. D., & Wang, R. (2006). *Journey to data quality*. MIT press.

Lee, Y. W., Strong, D. M., Kahn, B. K., & Wang, R. Y. (2002). AIMQ: a methodology for information quality assessment. *Information & Management*, *40*(2), 133–146.

Leuschner, R., Rogers, D. S., & Charvet, F. Ç. O. I. S. F. (2013). a Meta-Analysis of Supply Chain Integration and Firm Performance. *Journal of Supply Chain Management*, *49*(2), 34–57.

Leys, C., Klein, O., Dominicy, Y., & Ley, C. (2018). Detecting multivariate outliers: Use a robust variant of the Mahalanobis distance. *Journal of Experimental Social Psychology*, *74*(March 2017), 150–156.

Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, *49*(4), 764–766.

Li, L., Chi, T., Hao, T., & Yu, T. (2016). Customer demand analysis of the electronic commerce supply chain using Big Data. *Annals of Operations Research*. http://doi.org/10.1007/s10479-016-2342-x

Li, Q., Luo, H., Xie, P.-X., Feng, X.-Q., & Du, R.-Y. (2015). Product whole life-cycle and omni-channels data convergence oriented enterprise networks integration in a sensing environment. *Computers in Industry*, *70*, 23–45.

Li, Q. Y., Wang, L., & Xu, J. J. (2015). Production data analytics for production scheduling. *Industrial Engineering and Engineering Management (IEEM), 2015 IEEE International Conference On*, 1203–1207.

Li, Y., Ye, F., & Sheu, C. (2014). Social capital, information sharing and performance. *International Journal of Operations & Production Management*, *34*(11), 1440–1462.

Liao, S.-H., & Kuo, F.-I. (2014). The study of relationships between the collaboration for supply chain, supply chain capabilities and firm performance: A case of the Taiwanʹs TFT-LCD industry. *International Journal of Production Economics*, *156*, 295–304.

Liao, S. H., Wu, C. C., Hu, D. C., & Tsui, K. A. (2010). Relationships between knowledge acquisition, absorptive capacity and innovation capability: An empirical study on Taiwans financial and manufacturing industries. *Journal of Information Science*, *36*(1), 19–35.

Liao, Y., & Marsillac, E. (2015). External knowledge acquisition and innovation: the role of supply chain network-oriented flexibility and organisational awareness. *International Journal of Production Research*, *53*(18), 5437–5455.

Lii, P., & Kuo, F.-I. (2016). Innovation-oriented supply chain integration for combined competitiveness and firm performance. *International Journal of Production Economics*, *174*, 142–155.

Lin, H.-F. (2014). Understanding the determinants of electronic supply chain management system adoption: Using the technology–organization–environment framework. *Technological Forecasting and Social Change*, *86*, 80–92.

Linzer, D. A., & Lewis, J. B. (2011). poLCA: An R Package for Polytomous Variable Latent Class Analysis. *Journal of Statistical Software*, *42*(10), 1–29.

Little, J. D. C. (1961). A Proof for the Queuing Formula: L=λ W. *Operations Research*, *9*(3), 383–387. http://doi.org/10.1287/opre.9.3.383

Little, R. J. a, & Rubin, D. B. (2002). *Statistical Analysis with Missing Data*. *Statistical analysis with missing data Second edition*. http://doi.org/10.2307/1533221

Liu, C. L. E. (2012). An investigation of relationship learning in cross-border buyer-supplier relationships: The role of trust. *International Business Review*, *21*(3), 311–327.

Liu, G. (Jason), Shah, R., & Schroeder, R. G. (2012). The relationships among functional integration, mass customisation, and firm performance. *International Journal of Production Research*, *50*(3), 677–690.

Liu, H., Ke, W., Wei, K. K., & Hua, Z. (2013a). Effects of supply chain integration and market orientation on firm performance. *International Journal of Operations & Production Managemen*, *33*(3), 322–346.

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.

Liu, H., Ke, W., Wei, K. K., & Hua, Z. (2013b). 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.

Liu, K. (2014). Human Capital, Social Collaboration, and Patent Renewal Within U.S. Pharmaceutical Firms. *Journal of Management*, *40*(2), 616–636.

Liu, N., Roth, A. V, & Rabinovich, E. (2011). Antecedents and consequences of combinative competitive capabilities in manufacturing. *International Journal of Operations & Production Management*, *31*(11–12), 1250–1286.

Liu, X., Huang, Q., Dou, J., & Zhao, X. (2017). The impact of informal social interaction on innovation capability in the context of buyer-supplier dyads. *Journal of Business Research*, *78*, 314–322.

Liu, Y., Li, Z., Xiong, H., Gao, X., & Wu, J. (2010). Understanding of intenal clustering validation measures. *IEEE Internatinal Conference on Data Mining*, 911–916.

Loehlin, J. C., & Beaujean, A. A. (2017). *Latent Variable Models: An Introduction to Factor, Path, and Structural Equation Analysis*. *Routledge* (5th ed.). Routledge.

Lukman, T., Hackney, R., Popovič, A., Jaklič, J., & Irani, Z. (2011). Business Intelligence Maturity: The Economic Transitional Context Within Slovenia. *Information Systems Management*, *28*(3), 211–222.

Luzzini, D., Brandon-Jones, E., Brandon-Jones, A., & Spina, G. (2015). From sustainability commitment to performance: The role of intra- and inter-firm collaborative capabilities in the upstream supply chain. *International Journal of Production Economics*, *165*, 51–63.

Lynch, R. (2015). *Strategic Management* (7th ed.). Pearson Education Limited.

M Rocha, C., & Chelladurai, P. (2012). Item Parcels in Structural Equation Modeling: an

Applied Study in Sport Management. *International Journal of Psychology and Behavioral Sciences*, *2*(1), 46–53.

MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample Size in Factor Analysis. *Psychological Methods*, *4*(1), 84–99.

Mackelprang, A. W., Robinson, J. L., Bernardes, E., & Webb, G. S. (2014). The relationship between strategic supply chain integration and performance: A meta-analytic evaluation and implications for supply chain management research. *Journal of Business Logistics*, *35*(1), 71–96.

Mackenzie, S. B., Podsakoff, P. M., & Podsakoff, N. P. (2011). Construct measurement and validation procedures in MIS and behavioral research: Integrating new and existing techniques. *MIS Quarterly*, *35*(2), 293–334. Retrieved from http://dl.acm.org/citation.cfm?id=2017510%5Cnpapers3://publication/uuid/E4D3717C-7F3F-4791-8835-141D4309976B

MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G., & Sheets, V. (2002). A comparison of methods to test mediation and other intervening variable effects. *Psychological Methods*, *7*(1), 83–104.

Maital, S., & Seshadri, D. R. (2007). *Innovation, management : Strategies, concept and tools for growth and profit*. London: Sage Publications.

Malhotra, A., Gosain, S., & El Sawy, O. A. (2005). Absorptive capacity configurations in supply chains: Gearing for parter-enabled market knowledge creation. *MIS Quarterly*, *29*(1), 145–187.

Malhotra, M., & Grover, V. (1998). An assessment of survey research in POM: from constructs to theory. *Journal of Operations Management*, *16*(4), 407–425.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., … Byers. (2011). *Big data: The next frontier for innovation, competition, and productivity*. *McKinsey Global Institute*.

Markham, S. K., Kowolenko, M., & Michaelis, T. L. (2015). Unstructured Text Analytics to Support New Product Development Decisions. *Research Technology Management*, *58*(2), 30–39.

Marr, B. (2016). *Big Data in Practice (Use Cases) - How 45 Successful Companies Used Big Data Analytics to Deliver Extraordinary Results*. *John Wiley & Sons, Inc*.

Mata, F. J., Fuerst, W. L., & Barney, J. B. (1995). Information technology and sustained competitive advantage'. *MIS Quarterly, December*, *19*(4), 487–505.

Mauerhoefer, T., Strese, S., & Brettel, M. (2017). The Impact of Information Technology on New Product Development Performance. *Journal of Product Innovation Management*, *34*(6), 719–738.

Maulik, U., & Bandyopadhyay, S. (2002). Performance evaluation of some clustering algorithms and validity indices. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *24*(12), 1650–1654.

Mayring, P. (2003). *Qualitative Content Analysis* (8th ed.). Beltz Verlag, Weinheim, Germany.

Mazzola, E., Bruccoleri, M., & Perrone, G. (2015). Supply chain of innovation and new product development. *Journal of Purchasing and Supply Management*, *21*(4), 273–284.

McAfee, A., & Brynjolfsson, E. (2012). Big data: the management revolution. *Harvard Business Review*, *90*(10).

Meade, A. W., & Kroustalis, C. M. (2006). Problems With Item Parceling for of Measurement Invariance. *Organizational Research Methods*, *9*(3), 369–403.

Mehmood, R., Meriton, R., Graham, G., Hennelly, P., & Kumar, M. (2017). Exploring the influence of big data on city transport operations: a Markovian approach. *International Journal of Operations and Production Management*, *37*(1), 75–104.

Meinlschmidt, J., Foerstl, K., & Kirchoff, J. F. (2016). The role of absorptive and desorptive capacity (ACDC) in sustainable supply management A longitudinal analysis. *International Journal of Physical Distribution & Logistics Management*, *46*(2), 177–211.

Meixell, M., & Wu, S. (2001). Scenario analysis of demand in a technology market using leading indicators. *Semiconductor Manufacturing, IEEE …*, *14*(1), 65–75. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=909656

Mentzer, J. T., Keebler, J. S., Nix, N. W., Smith, C. D., & Zacharia, Z. G. (2001). Defining Supply Chain Management. *Journal of Business Logistics*, *22*(2), 1–25.

Meredith, J. R., Raturi, A., Amoako-Gympah, K., & Kaplan, B. (1989). Alternative research paradigms in operation. *Journal of Operations Management*, *8*(4), 297–326.

Meredith, R., Remington, S., O’Donnell, P., & Sharma, N. (2012). Organisational transformation through Business Intelligence: theory, the vendor perspective and a research agenda. *Journal of Decision Systems*, *21*(3), 187–201.

Merino, J., Caballero, I., Rivas, B., Serrano, M., & Piattini, M. (2016). A Data Quality in Use model for Big Data. *Future Generation Computer Systems*, *63*, 123–130.

Merschmann, U., & Thonemann, U. W. (2011). Supply chain flexibility, uncertainty and firm performance: An empirical analysis of German manufacturing firms. *International Journal of Production Economics*, *130*(1), 43–53.

Metan, G., Sabuncuoglu, I., & Pierreval, H. (2010). Real time selection of scheduling rules and knowledge extraction via dynamically controlled data mining. *International Journal of Production Research*, *48*(23), 6909–6938.

Milfont, T. L., & Fischer, R. (2010). Testing measurement invariance across groups : Applications in cross-. *International Journal of Psychological Research*, *3*(1), 111–121.

Millar, M. M., & Dillman, D. A. (2011). Improving response to web and mixed-mode surveys. *Public Opinion Quarterly*, *75*(2), 249–269.

Miller, J. W., Saldanha, J. P., Hunt, C. S., & Mello, J. E. (2013). Combining formal controls to improve firm performance. *Journal of Business Logistics*, *34*(4), 301–318.

Mingoti, S. A., & Lima, J. O. (2006). Comparing SOM neural network with Fuzzy c-means, K-means and traditional hierarchical clustering algorithms. *European Journal of Operational Research*, *174*(3), 1742–1759.

Mishra, D., Gunasekaran, A., Papadopoulos, T., Childe, S. J., & Oper Res, A. (2016). Big Data and supply chain management: a review and bibliometric analysis. *Annals of Operations Research*.

Mishra, N., & Singh, A. (2016). Use of twitter data for waste minimisation in beef supply chain. *Annals of Operations Research*.

Mithas, S., Ramasubbu, N., & Sambamurthy, V. (2011). How Information Managent Capability Influences Firm Performance. *MIS Quarterly*, *35*(1), 237–256. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=iih&AN=59550852&lang=pt-br&site=ehost-live

Mortenson, M. J. ., Doherty, N. F., & Robinson, S. (2015). Operational research from Taylorism to Terabytes: A research agenda for the analytic sage. *European Journal of Operational Research*, *241*(3), 583–595.

Mowery, D. C., & Oxley, J. E. (1995). Inward Technology Transfer and Competitiveness: The Role of National Innovation Systems. *Cambridge Journal of Economics*, *19*(1), 67–93.

Mowery, D. C., Oxley, J. E., & Silverman, B. . (1996). Strategic Alliances and Interfirm Knowledge Transfer. *Strategic Management Journal*, *17*(Winter Special Issue), 77–91.

Mukherjee, A., Mitchell, W., & Talbot, F. B. (2000). Impact of new manufacturing requirements on production line productivity and quality at a focused factory. *Journal of Operations Management*, *18*(2), 139–168.

Mulaik, S. A. (2010). *Foundations of Factor Analysis* (2nd ed.). Taylor & Francis Group, LLC.

Murovec, N., & Prodan, I. (2009). Absorptive capacity, its determinants, and influence on innovation output: Cross-cultural validation of the structural model. *Technovation*, *29*(12), 859–872. http://doi.org/10.1016/j.technovation.2009.05.010

Murtagh, F., & Legendre, P. (2014). Ward ' s Hierarchical Agglomerative Clustering Method : Which Algorithms Implement Ward' s Criterion ? *Journal of Classification*, *31*(October), 274–295. http://doi.org/10.1007/s00357-

Nada, O. A., ElMaraghy, H. A., & ElMaraghy, W. H. (2006). Quality prediction in manufacturing system design. *Journal of Manufacturing Systems*, *25*(3), 153–171.

Nagarur, N. (1992). Some performance measures of flexible manufacturing systems. *International Journal of Production Research*, *30*(4), 799–809.

Nagati, H., & Rebolledo, C. (2012). The role of relative absorptive capacity in improving suppliers' operational performance. *International Journal of Operations & Production Management*, *32*(5), 611–630.

Nahm, A. (2003). The impact of organizational structure on time-based manufacturing and plant performance. *Journal of Operations Management*, *21*(3), 281–306. 9

Najafi Tavani, S., Sharifi, H., & Ismail, H. S. (2013). A study of contingency relationships between supplier involvement, absorptive capacity and agile product innovation. *International Journal of Operations & Production Management*, *34*(1), 65–92.

Narasimhan, R., & Kim, S. W. (2002). Effect of supply chain integration on the relationship between diversification and performance: evidence from Japanese and Korean firms. *Journal of Operations Management*, *20*(3), 303–323.

Narasimhan, R., Kim, S. W., & Tan, K. C. (2008). An empirical investigation of supply chain strategy typologies and relationships to performance. *International Journal of Production Research*, *46*(18), 5231–5259.

Narasimhan, R., & Narayanan, S. (2013). Perspectives on Supply Network-Enabled Innovations. *Journal of Supply Chain Management*, *49*(4), 27–42.

Nassar, S. H. (2011). *Supply chain visibility and sustainable competitive advantage : An integrated model*. University of Bath.

Neaga, I., Liu, S., Xu, L., Chen, H., & Hao, Y. (2015). Cloud enabled big data business platform for logistics services: A research and development agenda. In B. Delibašić, E. J. Hernández, J. Papathanasiou, F. Dargam, P. Zaraté, R. Ribeiro, … I. Linden (Eds.), *Lecture Notes in Business Information Processing* (pp. 22–33). Cham: Springer International Publishing.

Neely, A., Gregory, M., & Platts, K. (1995). Performance measurement system design. *International Journal of Operations & Production Management*, *15*(4), 80–116.

Neely, A., Waggoner, D. B., D, & P. Kennerley, M. (1999). The forces that shape organisational performance measurement systems: *International Journal of Production Economics*, *60–61*, 53–60.

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

Neuman, W. L. (2014). *Social Research: Qualitative and Quantitative Approaches.* Pearson Education Limited.

Newsted, P., Huff, S., & Munro, M. (1998). Survey instruments in information systems. *MIS Quarterly: Management Information Systems*, *22*(4), 553–554.

Ng, S. C. H., Rungtusanatham, J. M., Zhao, X., & Ivanova, A. (2015). TQM and environmental uncertainty levels : profiles , fit , and firm performance. *International Journal of Production Research*, *7543*(June), 37–41.

Nita, S. (2015). Application of big data technology in support of food manufacturers' commodity demand forecasting. *NEC Technical Journal*, *10*(1), 90–93.

Noblet, J.-P., Simon, E., & Parent, R. (2011). Absorptive capacity: a proposed operationalization. *Knowledge Management Research & Practice*, *9*(4), 367–377.

Nunnally, J. C. (1978). *Psychometric Theory* (2nd ed.). New York: McGraw Hill.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). McGRAW-HILL, INC.

O'Donnell, P., Sipsma, S., & Watt, C. (2012). The critical issues facing business intelligence practitioners. *Journal of Decision Systems*, *21*(3), 203–216.

O'Leary-Kelly, S. W., & J. Vokurka, R. (1998). The empirical assessment of construct validity. *Journal of Operations Management*, *16*(4), 387–405.

Obayi, R., Koh, S. C., Oglethorpe, D., & Ebrahimi, S. M. (2017). Improving retail supply flexibility using buyer-supplier relational capabilities. *International Journal of Operations and Production Management*, *37*(3), 343–362.

OECD. (2001). *Understanding The Digital Divide*. Paris: OECD Publishing. Retrieved from http://dx.doi.org/10.1787/236405667766

OECD. (2014). *Measuring the Digital Economy: A new perspective*. OECD Publishing.

OECD. (2015). *Data-Driven Innovation: Big Data for Growth and Well-Being*. *OECD Digital Economy Papers* (Vol. 215). Paris: OECD Publishing.

Oh, S., Yang, H., & Kim, S. W. (2014). Managerial capabilities of information technology and firm performance: role of e-procurement system type. *International Journal of Production Research*, *52*(March 2015), 4488–4506.

Oke, A. (2013). Linking manufacturing flexibility to innovation performance in manufacturing plants. In *International Journal of Production Economics* (Vol. 143, pp.

242–247). Elsevier.

Oliveira, M. P. V. de, McCormack, K., & Trkman, P. (2012). Business analytics in supply chains – The contingent effect of business process maturity. *Expert Systems with Applications*, *39*(5), 5488–5498.

Olson, D. L. (2015). A Review of Supply Chain Data Mining Publications. *Journal of Supply Chain Management Science*, *xx*(xx), 1–13.

Olszak, C. M. (2016). Toward Better Understanding and Use of Business Intelligence in Organizations. *Information Systems Management*, *33*(2), 105–123.

Omachonu, V. K., Suthummanon, S., & Einspruch, N. G. (2004). The relationship between quality and quality cost for a manufacturing company. *International Journal of Quality and Reliability Management*, *21*(3), 277–290.

Opresnik, D., & Taisch, M. (2015). The value of big data in servitization. *International Journal of Production Economics*, *165*, 174–184.

Ou, C. S., Liu, F. C., Hung, Y. C., & Yen, D. C. (2010). A structural model of supply chain management on firm performance. *International Journal of Operations & Production Management*, *30*(5), 526–545.

Oxford Dictionary. (2016). In: Oxford Dictionary. Retrieved May 12, 2016, from http://www.oxforddictionaries.com/definition/english/data

Ozer, M. (2002). The role of flexibility in online business. *Business Horizons*, *45*(1), 61–69.

Pakhira, M. K., Bandyopadhyay, S., & Maulik, U. (2004). Validity index for crisp and fuzzy clusters. *Pattern Recognition*, *37*(3), 487–501.

Pallant, J. (2007). A step by step guide to data analysis using the SPSS program: SPSS survival manual. *Journal of Advanced Nursing*, *36*(3), 478–478.

Papadopoulos, T., Gunasekaran, A., Dubey, R., Altay, N., Childe, S. J., & Fosso- Wamba, S. (2016). The role of Big Data in explaining disaster resilience in supply chains for sustainability. *Journal of Cleaner Production*.

Park, H., Bellamy, M. A., & Basole, R. C. (2016). Visual analytics for supply network management: System design and evaluation. *Decision Support Systems*, *91*, 89–102. http://doi.org/http://dx.doi.org/10.1016/j.dss.2016.08.003

Patel, P. C., Terjesen, S., & Li, D. (2012). Enhancing effects of manufacturing flexibility through operational absorptive capacity and operational ambidexterity. *Journal of Operations Management*, *30*(3), 201–220. http://doi.org/10.1016/j.jom.2011.10.004

Patel, P. C., Terjesen, S., & Li, D. (2012). Enhancing effects of manufacturing flexibility through operational absorptive capacity and operational ambidexterity. *Journal of Operations Management*, *30*(3), 201–220. http://doi.org/10.1016/j.jom.2011.10.004

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. http://doi.org/10.1287/isre.1060.0094

Pearl, J. (2012). The causal foundations of structural equation modeling. In *Handbook of Structural Equation Modeling* (pp. 68–91).

Peng, D. X., Schroeder, R. G., & Shah, R. (2008). Linking routines to operations capabilities: A new perspective. *Journal of Operations Management*, *26*(6), 730–748.

Penrose, E. T. (1959). *The Theory of the Growth of the Firm*. *London: Blackwells*. New

York: John Wiley & Sons.

Peppard, J., & Ward, J. (2004). Beyond strategic information systems: Towards an IS capability. *Journal of Strategic Information Systems, 13*(2), 167–194.

Persson, O. (2010). Bibexcel–a toolbox for bibliometricians.

Peteraf, M. A. (1993). The cornerstones of competitive advantage: A resource-based view. *Strategic Management Journal*, *14*(3), 179–191.

Petroni, A., & Panciroli, B. (2002). Innovation as a determinant of suppliers' roles and performances: an empirical study in the food machinery industry. *European Journal of Purchasing & Supply Management, 8*(3), 135–149.

Pfohl, H., & Buse, H. P. (2000). Inter-organizational logistics systems in flexible production networks: An organizational capabilities perspective. *International Journal of Physical Distribution & Logistics Management*, *30*(5), 388–408.

Phan, A. C., Abdallah, A. B., & Matsui, Y. (2011). Quality management practices and competitive performance: Empirical evidence from Japanese manufacturing companies. *International Journal of Production Economics*, *133*(2), 518–529.

Pihlajamaa, M., Kaipia, R., Säilä, J., & Tanskanen, K. (2017). Can supplier innovations substitute for internal R&D? A multiple case study from an absorptive capacity perspective. *Journal of Purchasing and Supply Management*, *23*(4), 242–255. http://doi.org/10.1016/j.pursup.2017.08.002

Piller, G., & Hagedorn, J. (2011). Business Benefits And Application Capabilities Enabled By In-Memory Data Management. *Proceedings Des Workshops Innovative Unternehmensanwendungen Mit In-Memory Data Management,* 45–56. Retrieved from http://subs.emis.de/LNI/Proceedings/Proceedings193/45.pdf

Pinheiro De Lima, E., Gouvea Da Costa, S. E., Angelis, J. J., & Munik, J. (2013). Performance measurement systems: A consensual analysis of their roles. *International Journal of Production Economics*, *146*(2), 524–542.

Pinsonneault, A., & Kraemer, K. L. (1993). Survey research methodology in management information systems: An assessment. *Journal of Management Information Systems*, *10*(2), 75–105.

Pipino, L. L., Lee, Y. W., Wang, R. Y., Lowell Yang Lee, M. W., & Yang, R. Y. (2002). Data Quality Assessment. *Communications of the ACM, 45*(4), 211.

Pittaway, L., Robertson, M., Munir, K., Denyer, D., & Neely, A. (2004). Networking and innovation: a systematic review of the evidence. *International Journal of Management Reviews*, *5–6*(3–4), 137–168.

Pituch, K. A., & Stevens, J. P. (2016). *Applied Multivariate Statistics for the Social Sciences*. *Routledge*.

Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common Method Biases in Behavioral Research: A Critical Review of the Literature and Recommended Remedies. *Journal of Applied Psychology*, *88*(5), 879–903.

Podsakoff, P. M. P. M. M., & Organ, D. W. (1986). Self-Reports in Organizational Research: Problems and Prospects. *Journal of Management, 12*(4), 531–544.

Popovič, A., Coelho, P. S., & Jurij Jaklič. (2009). The impact of business intelligence system maturity on information quality. *Information Research*, *14*(4), paper 417. Retrieved from http://informationr.net/ir/14-4/paper417.html

Popovič, A., Hackney, R., Coelho, P. S., & Jaklič, J. (2012). Towards business intelligence systems success: Effects of maturity and culture on analytical decision making. *Decision Support Systems*, *54*(1), 729–739.

Porter, A. (2009). *Operations Management*.

Porter, M. E. (1985). Competitive Advantage: Creating and sustaining superior performance. New York: The free press.

Porter, M. E. (1985). *Competitive Strategy: The Core Concepts*. *Creating and Sustaining Competitive Advantage*. http://doi.org/10.1007/978-3-319-54540-0

Powell, T., & Dent-Micallef, A. (1997). Information Technology as Competitive Advantage : The Role of Human , Business , and Technology Resources. *Strategic Management Journal*, *18*(5), 375–405.

Preacher, K. J., & Hayes, A. F. (2008a). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods*, *40*(3), 879–891.

Preacher, K. J., & Hayes, A. F. (2008b). Contemporary approaches to assessing mediation in communication research. *The Sage Sourcebook of Advanced Data Analysis Methods for Communication Research.*, 13–54.

Prescott, M. (2016). Big Data: Innovation and Competitive Advantage in an Information Media Analytics Company. *Journal of Innovation Management*, *1*, 92–113. Retrieved from http://www.open-jim.org/article/view/152

Preston, D. S., Chen, D. Q., Swink, M., & Meade, L. (2017). Generating Supplier Benefits through Buyer-Enabled Knowledge Enrichment: A Social Capital Perspective. *Decision Sciences*, *48*(2), 248–287.

Priem, R. L., & Butler, J. E. (2001). Is the Resource-Based " View " a Useful Perspective for Strategic Management Research ? *The Academy of Management Review*, *26*(1), 22–40.

Puget, J. F. (2015). The Analytics Maturity Model. Retrieved August 4, 2016, from https://www.ibm.com/developerworks/community/blogs/jfp/entry/the_analytics_maturity_model?lang=en

Punj, G., & Stewart, D. W. (1995). Cluster Analysis in Marketing Research: Review and Suggestions for Application. *Journal of Marketing Research*, *20*(2), 134–148.

Qrunfleh, S., & Tarafdar, M. (2013). Lean and agile supply chain strategies and supply chain responsiveness: the role of strategic supplier partnership and postponement. *Supply Chain Management: An International Journal*, *18*(6), 571–582.

Rada, V. D. de, & Domínguez-Álvarez, J. A. (2014). Response Quality of Self-Administered Questionnaires: A Comparison Between Paper and Web Questionnaires. *Social Science Computer Review*, *32*(2), 256–269.

Radcliffe, J. (2014). Leverage a Big Data Maturity Model to Build Your Big Data Roadmap. *Radcliffe Advisory Services Ltd*, (January). Retrieved from www.radcliffeadvisory.com

Radke, A. M., & Tseng, M. M. (2015). Design Considerations for Building Distributed Supply Chain Management Systems Based on Cloud Computing. *Journal of Manufacturing Science and Engineering*, *137*(4), 40906.

Raghunathan, S. (1999). Impact of information quality and decision-maker quality on decision quality: A theoretical model and simulation analysis. *Decision Support Systems*, *26*(4), 275–286.

Rai, Patnayakuni, & Seth. (2006). Firm Performance Impacts of Digitally Enabled Supply Chain Integration Capabilities. *MIS Quarterly*, *30*(2), 225.

Ramsey, S. R., Thompson, K. L., McKenzie, M., & Rosenbaum, A. (2016). Psychological research in the internet age: The quality of web-based data. *Computers in Human Behavior*, *58*, 354–360.

Rea, L. M., & Parker, R. A. (2014). *Designing and conducting survey research: a comprehensive guide*. *John Wiley & Sons, Inc.*

Redman, T. C. (1996). *Data Quality for the Information Age*.

Redman, T. C. (1998). The impact of poor data quality on the typical enterprise. *Communications of the ACM*, *41*(2), 79–82.

Redman, T. C. (2001). *Data Quality: The Field Guide*. Retrieved from http://books.google.com.br/books?id=iE2qsxDyEEYC

Redman, T. C. (2008). *Data Driven: Profiting from Your Most Important Business Asset*. *Harvard Business Press*. Boston.

Reed, R., & Defillippi, R. J. (1990). Casual Ambiguity, Barriers to Immitation and Sutainable Competitive Advantage. *The Academy of Management Review*.

Rehman, M. H. ur, Chang, V., Batool, A., & Wah, T. Y. (2016). Big data reduction framework for value creation in sustainable enterprises. *International Journal of Information Management*, *36*(6, Part A), 917–928.

Revilla, E., Sáenz, M. J., & Knoppen, D. (2013). Towards an empirical typology of buyer-supplier relationships based on absorptive capacity. *International Journal of Production Research*, *51*(10), 2935–2951.

Richey Jr, R. G., Morgan, T. R., Lindsey-Hall, K., & Adams, F. G. (2016). A global exploration of Big Data in the supply chain. *International Journal of Physical Distribution & Logistics Management*, *46*(8), 710–739.

Riikkinen, R., Kauppi, K., & Salmi, A. (2016). Learning Sustainability? Absorptive capacities as drivers of sustainability in MNCs’ purchasing. *International Business Review*, *26*(6), 1075–1087.

Roberts, N., Galluch, P. S., Dinger, M., & Grover, V. (2012). Absorptive Capacity and Information Systems Research: Review, Synthesis, and Directions for Future Research. *MIS Quarterly*, *36*(2), 625–648.

Robey, D. (1996). Research Commentary: Diversity in Information Systems Research: Threat, Promise, and Responsibility. *Information Systems Research*, *7*(4), 400–408.

Rogers, P. P., Ojha, D., & White, R. E. (2011). Conceptualising complementarities in manufacturing flexibility: A comprehensive view. *International Journal of Production Research*, *49*(12), 3767–3793.

Roldan Bravo, M. I., Ruiz Moreno, A., & Llorens-Montes, F. J. (2016). Supply network-enabled innovations. An analysis based on dependence and complementarity of capabilities. *Supply Chain Management*, *21*(5), 642–660.

Rönkkö, M., McIntosh, C. N., Antonakis, J., & Edwards, J. R. (2016). Partial least squares path modeling: Time for some serious second thoughts. *Journal of Operations Management*, *47–48*, 9–27.

Rosenzweig, E. D., Roth, A. V., & Dean, J. W. (2003). The influence of an integration strategy on competitive capabilities and business performance: An exploratory study of consumer products manufacturers. *Journal of Operations Management*, *21*(4), 437–456.

Rothaermel, F. T. (2017). *Strategic Management* (3rd ed.). McGraw-Hill Education.

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, *20*, 53–65.

Rozados, V. I., & Tjahjono, B. (2014). Big Data Analytics in Supply Chain Management : Trends and Related Research. In *6th International Conference on Operations and Supply Chain Management, Bali*.

Rungtusanatham, M. J., Choi, T. Y., Hollingworth, D. G., Wu, Z., & Forza, C. (2003). Survey research in operations management: Historical analyses. *Journal of Operations Management*, *21*(4), 475–488.

Ryu, K., Park, J., & Park, J. (2006). A Data Quality Management Maturity Model. *ETRI Journal*, *28*(2), 191–204.

Saenz, M. J., Revilla, E., & Knoppen, D. (2014). Absorptive capacity in buyer-supplier relationships: Empirical evidence of its mediating role. *Journal of Supply Chain Management*, *50*(2), 18–40.

Sáenz, M. J., Revilla, E., & Knoppen, D. (2014). Absorptive capacity in buyer-supplier relationships: Empirical evidence of its mediating role. *Journal of Supply Chain Management*, *50*(2), 18–40.

Sahay, B. S., & Ranjan, J. (2008). Real time business intelligence in supply chain analytics. *Information Management & Computer Security*, *16*(1), 28–48.

Sajadfar, N., & Ma, Y. (2015). A hybrid cost estimation framework based on feature-oriented data mining approach. *Advanced Engineering Informatics*, *29*(3), 633–647.

Sambamurthy, Bharadwaj, & Grover. (2003). Shaping Agility through Digital Options: Reconceptualizing the Role of Information Technology in Contemporary Firms. *MIS Quarterly*, *27*(2), 237.

Sánchez, A., & Pérez, M. (2005). Supply chain flexibility and firm performance: A conceptual model and empirical study in the automotive industry. *International Journal of Operations & Production Management*, *25*(7), 681–700.

Sanders, N. (2014). *Big Data Driven Supply Chain Management: A Framework for Implementing Analytics and Turning Information Into Intelligence*. Pearson Education, Inc. Retrieved from http://books.google.com.au/books?hl=de&lr=&id=-b2LAwAAQBAJ&oi=fnd&pg=PR6&dq=customer+driven+supply+chain&ots=RNP3I_URAY&sig=GO6InXDjMpl95Upqa-_Rqvmz-cQ

Sanders, N. R. (2016). How to Use Big Data to Drive Your Supply Chain. *California Management Review*, *58*(3), 26–48.

Sangari, M. S., & Razmi, J. (2015). Business intelligence competence, agile capabilities, and agile performance in supply chain: An empirical study. *The International Journal of Logistics Management*, *26*(2), 356–380.

Santiago Rivera, D., & Shanks, G. (2015). A Dashboard to Support Management of Business Analytics Capabilities. *Journal of Decision Systems*, *24*(1), 73–86.

Saraçli, S., Doğan, N., & Doğan, I. (2013). Comparison of hierarchical cluster analysis methods by cophenetic correlation. *Journal of Inequalities and Applications*, 1–8.

Saris, W. E., & Gallhofer, I. N. (2014). *Design, Evaluation, and Analysis of Questionnaires* *for Survey Research* (2nd ed.). John Wiley & Sons, Inc.

SAS. (2013). *Big Data Analytics Adoption and Employment Trends, 2012-2017*. London. Retrieved from http://www.sas.com/offices/europe/uk/downloads/bigdata/eskills/eskills.pdf

Sattler, M. (2011). *Excellence in Innovation Management: A Meta-Analytic Review on the Predictors of Innovation Performance*. GABLER RESEARCH.

Saunders, M., Lewis, P., & Thornhill, A. (2009). *Research Methods for Business Students*. *Pearson Education Limited* (5th ed.).

Saunders, M., Lewis, P., & Thornhill, A. (2016). *Research Methods For Business Students*. *Pearson Education UK* (7th ed.). Pearson Education Limited.

Savitskie, K. (2007). Internal and external logistics information technologies: The performance impact in an international setting. *International Journal of Physical Distribution & Logistics Management*, *37*(6), 454–468.

Scherrer-Rathje, M., Deflorin, P., & Anand, G. (2014). Manufacturing flexibility through outsourcing: effects of contingencies. *International Journal of Operations & Production Management*, *34*(9), 1210–1242.

Scherrer-Rathje, M., Deflorin, P., & Anand, G. (2014). Manufacturing flexibility through outsourcing: effects of contingencies. *International Journal of Operations & Production Management*, *34*(9), 1210–1242.

Schiele, H. (2007). Supply-management maturity, cost savings and purchasing absorptive capacity: Testing the procurement-performance link. *Journal of Purchasing and Supply Management*, *13*(4), 274–293.

Schildt, H., Keil, T., & Maula, M. (2012). The temporal effects of relative and firm-level absorptive capacity on interorganizational learning. *Strategic Management Journal*, *33*(10), 1154–1173.

Schmenner, R. W., Wassenhove, L. Van, Ketokivi, M., Heyl, J., & Lusch, R. F. (2009). Too much theory, not enough understanding. *Journal of Operations Management*, *27*(5), 339–343.

Schmidt, B., Flannery, P., & DeSantis, M. (2014). Real-Time Predictive Analytics, Big Data & Energy Market Efficiency: Key to Efficient Markets and Lower Prices for Consumers. *Applied Mechanics and Materials*, *704*(Mechanics and Control Engineering III), 453–458.

Schminkey, D. L., von Oertzen, T., & Bullock, L. (2016). Handling Missing Data With Multilevel Structural Equation Modeling and Full Information Maximum Likelihood Techniques. *Research in Nursing & Health*, *39*(4), 286–297.

Schoenherr, T., Griffith, D. A., & Chandra, A. (2014). Knowledge management in supply chains: The role of explicit and tacit knowledge. *Journal of Business Logistics*, *35*(2), 121–135.

Schoenherr, T., & Speier-Pero, C. (2015). Data Science, Predictive Analytics, and Big Data in Supply Chain Management: Current State and Future Potential. *Journal of Business Logistics*, *36*(1), 120–132.

Schreiber, J. B., Stage, F. K., King, J., Nora, A., & Barlow, E. A. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. *Journal of Educational Research*, *99*(6), 323–337.

Schumacker, R. E., & Lomax, R. G. (2010). *A Beginner's Guide to structural equation*

Modeling. http://doi.org/10.1002/9781118133880.hop202023

Schwab, D. (1980). Construct validity in organizational behavior. *Research in Organizational Behavior*, *2*(1), 3–43. Retrieved from http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Construct+validity+in+organizational+behavior#0%5Cnftp://209-6-206-114.c3-0.smr-ubr2.sbo-smr.ma.cable.rcn.com/Seagate_Backup_Plus_Drive/JOTA/MIT/11.233 Research Design/Readings/Construct valid

Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, *6*(2), 461–464. http://doi.org/10.1214/aos/1176344136

Seah, M., Hsieh, M. H., & Weng, P.-D. (2010). A case analysis of Savecom: The role of indigenous leadership in implementing a business intelligence system. *International Journal of Information Management*, *30*(4), 368–373.

See-To, E. W. K., & Ngai, E. W. T. (2016). Customer reviews for demand distribution and sales nowcasting: a big data approach. *Annals of Operations Research*, 1–17.

Sekaran, U. (2003). *Research Methods for Business - A Skill Building Approach*. *Journal of Chemical Information and Modeling* (4th ed., Vol. 53). John Wiley & Sons, Inc.

Selvage, M. Y., Judah, S., & Jain, A. (2017). *Magic Quadrant for Data Quality Tools*. *Gartner*.

Seo, Y.-J. ., Dinwoodie, J. ., & Kwak, D.-W. . (2014). The impact of innovativeness on supply chain performance: is supply chain integration a missing link? *Supply Chain Management*, *19*, 733–746.

Seog-Chan, O., & Hildreth, A. J. (2016). *Analytics for Smart Energy Management:Tools and Applications for Sustainable Manufacturing*. Springer.

Sethi, A. K., & Sethi, S. P. (1990). Flexibility in manufacturing: A survey. *International Journal of Flexible Manufacturing Systems*, *2*(4), 289–328.

Setia, P., & Patel, P. C. (2013). How information systems help create OM capabilities: Consequents and antecedents of operational absorptive capacity. *Journal of Operations Management*, *31*(6), 409–431.

Setia, P., Venkatesh, V., & Joglekar, S. (2013). Leveraging Digital Technologies: How Information Quality Leads to Localized Capabilities and Customer Service Performance. *MIS Quarterly*, *37*(2), 565–590.

Seuring, S., & Müller, M. (2008). From a literature review to a conceptual framework for sustainable supply chain management. *Journal of Cleaner Production*, *16*(15), 1699–1710.

Shen, C., Riaz, Z., Palle, M. S., Jin, Q., & Peña-Mora, F. (2015). Data Quality Assurance in International Supply Chains: An Application of the Value Cycle Approach. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *9373*, 247–260.

Shen, W., Hao, Q., Mak, H., Neelamkavil, J., Xie, H., Dickinson, J., … Xue, H. (2010). Systems integration and collaboration in architecture, engineering, construction, and facilities management: A review. *Advanced Engineering Informatics*, *24*(2), 196–207.

Shepherd, C., & Gunter, H. (2006). *Measuring supply chain performance: current research and future directions*. *International Journal of Productivity and Performance Management* (Vol. 55).

Sherer, S. A. (2005). From supply chain management to value network advocacy:

implications for e-supply chains. *Supply Chain Management: An International Journal*, *10*(2), 77–83.

Shetty, Y. K. (1987). Product Quality and Competitive Strategy, Business Horizons, 46-52.

Shi, D., & Daniels, R. L. (2003). A survey of manufacturing flexibility: Implications for e-business flexibility. *IBM Systems Journal*, *42*(3), 414–427.

Shim, J. P., Warkentin, M., Courtney, J. F., Power, D. J., Sharda, R., & Carlsson, C. (2002). Past, present, and future of decision support technology. *Decision Support Systems*, *33*(2), 111–126.

Shin, B., & Kim, G. (2011). Investigating the reliability of second-order formative measurement in information systems research. *European Journal of Information Systems*, *20*(February), 608–623.

Shin, S. J., Woo, J., & Rachuri, S. (2014). Predictive analytics model for power consumption in manufacturing. *Procedia CIRP*, *15*, 153–158. http://doi.org/10.1016/j.procir.2014.06.036

Sidi, F., Shariat Panahy, P. H., Affendey, L. S., Jabar, M. a, Ibrahim, H., & Mustapha, a. (2012). Data quality: A survey of data quality dimensions. *Information Retrieval & Knowledge Management (CAMP), 2012 International Conference On,* 300–304.

Singh, P. J., & Power, D. (2009). The nature and effectiveness of collaboration between firms, their customers and suppliers: a supply chain perspective. *Supply Chain Management: An International Journal*, *14*(3), 189–200.

Singh, P. J., & Power, D. (2014). Innovative knowledge sharing, supply chain integration and firm performance of Australian manufacturing firms. *International Journal of Production Research*, *52*(21), 6416–6433.

Sirmon, D. G., Hitt, M. A., & Ireland, R. D. (2007). Managing Firm Resources In Dynamic Environments To Create Value: Looking Inside The Black Box. *The Academy of Management Review*, *32*(1), 273–292.

Sivakumar, K., & Roy, S. (2004). Knowledge redundancy in supply chains: a framework. *Supply Chain Management-An International Journal*, *9*(3–4), 241–249.

Sivo, S. A., Xitao, F. A. N., Witta, E. L., & Willse, J. T. (2006). The search for “optimal” cutoff properties: Fit index criteria in structural equation modeling. *Journal of Experimental Education*, *74*(3), 267–288.

Slack, N., Brandon-Jones, A., & Johnston, R. (2016). *Operations Management* (8th ed.). London, UK.

Slack, N., Chambers, S., & Johnston, R. (2007). *Operations Management*. *Operations Management* (5th ed.). Prentice-Hall, Inc. http://doi.org/9780132342711

So, K. C., & Song, J. S. (1998). Price, delivery time guarantees and capacity selection. *European Journal of Operational Research*, *111*(1), 28–49.

Soban, D., Thornhill, D., Salunkhe, S., & Long, A. (2016). Visual Analytics as an Enabler for Manufacturing Process Decision-making. *Procedia CIRP*, *56*, 209–214.

Song, M., & Thieme, J. (2009). The role of suppliers in market intelligence gathering for radical and incremental innovation. *Journal of Product Innovation Management*, *26*(1), 43–57.

Song, Z., Sun, Y., Wan, J., & Liang, P. (2017). Data quality management for service-oriented manufacturing cyber-physical systems. *Computers and Electrical*

Engineering, *64*, 1339–1351.

Sonka, S. (2014). Big data and the ag sector: More than lots of numbers. *International Food and Agribusiness Management Review*, *17*(1), 1–20.

Souza, G. C. (2014). Supply chain analytics. *Business Horizons*, *57*(5), 595–605. http://doi.org/10.1016/j.bushor.2014.06.004

Spanner, G. E., Nuno, J. P., & Chandra, C. (1993). Time-based strategies - Theory and practice. *Long Range Planning*, *26*(4), 91.

Spruit, M., & Sacu, C. (2015). DWCMM: The Data Warehouse Capability Maturity Model. *Journal of Universal Computer Science*, *21*(11), 1508–1534. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.227.4569&rep=rep1&type=pdf

Stalk Jr., G., & Hout, T. M. (1990). *Competing Against Time*.

Stefanovic, N. (2014). Proactive Supply Chain Performance Management with Predictive Analytics. *The Scientific World Journal*, 1–17.

Stefanovic, N. (2015). Collaborative Predictive Business Intelligence Model for Spare Parts Inventory Replenishment. *Computer Science & Information Systems*, *12*(3), 911.

Straub, D. W. (1989). Validating Instruments in MIS Research. *MIS Quarterly*, *13*(2), 147.

Strong, D. M. (1997). IT process designs for improving information quality and reducing exception handling: A simulation experiment. *Information and Management*, *31*(5), 251–263.

Sue, V. M., & Ritter, L. A. (2007). *Conducting Online Surveys*. London: Sage Publications, Inc.

Sulaiman, H., Cob, Z. C., & Ali, N. (2015). Big Data Maturity Model for Malaysian Zakat Institutions to Embark on Big Data Initiatives. *4th International Conference on Software Engineering and Computer Systems (ICSECS)*, 61–66.

Sundbo, J., & Fuglsang, L. (2002). *Innovation as strategic reflexivity*. *Routledge advances in management and business studies 22*. London, UK: Routledge.

Tabachnick, B. G., & Fidell, L. S. (2012). *Using Multivariate Statistics*. *Pearson Education Limited* (6th ed.).

Tachizawa, E. M., Alvarez-Gil, M. J., & Montes-Sancho, M. J. (2015). How “smart cities” will change supply chain management. *Supply Chain Management: An International Journal*, *20*(3), 237–248.

Takeishi, A. (2001). Bridging intra- and inter-firm boundaries: management of supplier involvement in automobile product development. *Strategic Management Journal*, *22*(November 1998), 403–433. Retrieved from http://onlinelibrary.wiley.com/doi/10.1002/smj.164/abstract

Tan, K. C., Kannan, V. R., Hsu, C.-C., & Leong, G. K. (2010). Supply chain information and relational alignments: mediators of EDI on firm performance. *International Journal of Physical Distribution & Logistics Management*, *40*(5), 377–394.

Tan, K. H., Zhan, Y., Ji, G., Ye, F., & Chang, C. (2015). Harvesting big data to enhance supply chain innovation capabilities: An analytic infrastructure based on deduction graph. *International Journal of Production Economics*, *165*, 223–233.

Tanriverdi. (2005). Information Technology Relatedness, Knowledge Management

Capability, and Performance of Multibusiness Firms. *MIS Quarterly*, *29*(2), 311.

Tao, Y., & Storey, V. C. (2015). A Recommender System for Two-sided Markets : Understanding Customers Sentiment in Social Networks. *Twenty-First Americas Conference on Information Systems*, 1–11.

Tate, W. L., Ellram, L. M., & Dooley, K. J. (2014). The impact of transaction costs and institutional pressure on supplier environmental practices. *International Journal of Physical Distribution & Logistics Management*, *44*(5), 353–372.

Tavani, S. N., Sharifi, H., Soleimanof, S., & Najmi, M. (2013). An empirical study of firm's absorptive capacity dimensions, supplier involvement and new product development performance. *International Journal of Production Research*, *51*(11), 3385–3403.

Tee, S. W., Bowen, P. L., Doyle, P., & Rohde, F. H. (2007). Factors influencing organizations to improve data quality in their information systems. *Accounting and Finance*, *47*(3), 335–355.

Teece, D. J., Pisano, G., & Shuen, A. (1997). Dynamic Capabilities and Strategic Management. *Strategic Management Journal*, *18*(7), 509–533. Retrieved from http://www.jstor.org/stable/3088148

The Economist. (2010). The data deluge. Retrieved from http://www.economist.com/node/15579717

Thompson, B. (2004). *Exploratory and confirmatory factor analysis: Understanding concepts and applications*. Washington, DC, USA: American Psychological Association.

Thompson, V. A. (1965). Bureaucracy and Innovation. *Administrative Science Quarterly*, *10*(1), 1–20.

Tien, J. M. (2012). The next industrial revolution: Integrated services and goods. *Journal of Systems Science and Systems Engineering*, *21*(3), 257–296.

Tien, J. M. (2015). Internet of connected ServGoods: Considerations, consequences and concerns. *Journal of Systems Science and Systems Engineering*, *24*(2), 130–167.

Tiep, N. D. (2006). Building external manufacturing capability in emerging markets: Honda's knowledge transfer and the role of local suppliers' responsiveness. *Journal of Asia-Pacific Business*, *7*(4), 77–95.

Ting, S. L., Tse, Y. K., Ho, G. T. S., Chung, S. H., & Pang, G. (2014). Mining logistics data to assure the quality in a sustainable food supply chain: A case in the red wine industry. *International Journal of Production Economics*, *152*, 200–209.

Todd, L. (2013). *The Oxford Handbook of Quantitative Methods, Volume 2 Statistical analysis*. *Oxford Library of Psychology*. Oxford University Press.

Todorova, G., & Durisin, B. (2007). Absorptive Capacity : Valuing a Reconceptualization. *The Academy of Management Review*, *32*(3), 774–786.

Tomar, G. S., Chaudhari, N. S., Bhadoria, R. S., & Deka, G. C. (Eds.). (2017). *The human element of Big Data: issues, analytics, and performance*. Taylor & Francis Group, LLC.

Tomarken, A. J., & Waller, N. G. (2005). Structural Equation Modeling: Strengths, Limitations, and Misconceptions. *Annual Review of Clinical Psychology*, *1*(1), 31–65.

Tranfield, D., Denyer, D., & Smart, P. (2003). Towards a Methodology for Developing Evidence-Informed Management Knowledge by Means of Systematic Review. *British*

Journal of Management, *14*(3), 207–222.

Treiblmaier, H., Bentler, P. M., & Mair, P. (2011). Formative Constructs Implemented via Common Factors. *Structural Equation Modeling: A Multidisciplinary Journal*, *18*(1), 1–17.

Trkman, P., McCormack, K., de Oliveira, M. P. V., & Ladeira, M. B. (2010). The impact of business analytics on supply chain performance. *Decision Support Systems*, *49*(3), 318–327.

Tsai, C.-W., Lai, C.-F., Chao, H.-C., & Vasilakos, A. V. (2015). Big data analytics: a survey. *Journal of Big Data*, *2*(1), 21.

Tseng, P.-H., & Liao, C.-H. (2015). Supply chain integration, information technology, market orientation and firm performance in container shipping firms. *The International Journal of Logistics Management*, *26*(1), 82–106.

Tu, Q., Vonderembse, M. A., Ragu-Nathan, T. S., & Sharkey, T. W. (2006). Absorptive capacity: Enhancing the assimilation of time-based manufacturing practices. *Journal of Operations Management*, *24*(5), 692–710.

Tuarob, S., & Tucker, C. S. (2015). Quantifying Product Favorability and Extracting Notable Product Features Using Large Scale Social Media Data. *Journal of Computing and Information Science in Engineering*, *15*(3), 031003.

Tzokas, N., Kim, Y. A., Akbar, H., & Al-Dajani, H. (2015). Absorptive capacity and performance: The role of customer relationship and technological capabilities in high-tech SMEs. *Industrial Marketing Management*, *47*, 134–142.

UK Government. (2014). *High value manufacturing (2012-2015)*. Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/362294/High_Value_Manufacturing_Strategy_2012-15.pdf

van der Spoel, S., Amrit, C., & van Hillegersberg, J. (2015). Predictive analytics for truck arrival time estimation: a field study at a European distribution center. *International Journal of Production Research*, (May), 1–17.

Vehovar, V., Sicherl, P., Hüsing, T., Dolničar, V., & Dolnicar, V. (2006). Methodological Challenges of Digital Divide Measurements. *The Information Society*, *22*(5), 279–290.

Vera-Baquero, A., Colomo Palacios, R., Stantchev, V., & Molloy, O. (2015). Leveraging big-data for business process analytics. *The Learning Organization*, *22*(4), 215–228.

Vickery, S. K. K., Droge, C., Setia, P., & Sambamurthy, V. (2010). Supply chain information technologies and organisational initiatives: complementary versus independent effects on agility and firm performance. *International Journal of Production Research*, *48*(23), 7025–7042.

Vokurka, R. J., & O'Leary-Kelly, S. W. (2000a). A review of empirical research on manufacturing flexibility. *Journal of Operations Management*, *18*(4), 485–501.

Vokurka, R. J., & O'Leary-Kelly, S. W. (2000b). A review of empirical research on manufacturing flexibility. *Journal of Operations Management*, *18*(4), 485–501.

vom Brocke, J., Debortoli, S., Müller, O., & Reuter, N. (2014). How in-memory technology can create business value: Insights from the hilti case. *Communications of the Association for Information Systems*, *34*(1), 151–168.

Wacker, J. (1998). A definition of theory: research guidelines for different theory-building research methods in operations management. *Journal of Operations Management*, *16*(4), 361–385.

Wade, M., & Hulland, J. (2004). Review: The resource-based view and information systems research: Review, extension, and suggestions for future research. *MIS Quarterly: Management Information Systems*, *28*(1), 107–142.

Wagner, S. M. (2006). A firm's responses to deficient suppliers and competitive advantage. *Journal of Business Research, 59*(6), 686–695.

Walesiak, M. (1999). Distance measure for ordinal data. *Argumenta Oeconomica*, *2*(8), 167–173.

Wall, M. (2003). Manufacturing flexibility. *Automotive Industries*, *183*(10), 44. http://doi.org/10.1111/j.1365-2591.2007.01265.x

Waller, M. A., & Fawcett, S. E. (2013a). Click Here for a Data Scientist: Big Data, Predictive Analytics, and Theory Development in the Era of a Maker Movement Supply Chain. *Journal of Business Logistics*, *34*(4), 249–252.

Waller, M. A., & Fawcett, S. E. (2013b). Data Science, Predictive Analytics, and Big Data: A Revolution That Will Transform Supply Chain Design and Management. *Journal of Business Logistics*, *34*(2), 77–84.

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.

Wamba, S. F., Gunasekaran, A., Akter, S., Ren, S. J., Dubey, R., & Childe, S. J. (2017). Big data analytics and firm performance: Effects of dynamic capabilities. *Journal of Business Research*, *20*, 356–365.

Wand, Y., & Wang, R. Y. (1996). Anchoring data quality dimensions in ontological foundations. *Communications of the ACM*, *39*(11), 86–95.

Wang, E. T. G., Chou, F. K. Y., Lee, N. C. A., & Lai, S. Z. (2014). Can intrafirm IT skills benefit interfirm integration and performance? *Information and Management*, *51*(7), 924–938.

Wang, G., Gunasekaran, A., Ngai, E. W. T., & Papadopoulos, T. (2016). Big data analytics in logistics and supply chain management: Certain investigations for research and applications. *International Journal of Production Economics*, *176*, 98–110.

Wang, L., Yeung, J. H. Y., & Zhang, M. (2011). The impact of trust and contract on innovation performance: The moderating role of environmental uncertainty. *International Journal of Production Economics*, *134*(1), 114–122.

Wang, N., Liang, H., Jia, Y., Ge, S., Xue, Y., & Wang, Z. (2015). Cloud computing research in the IS discipline: A citation/co-citation analysis. *Decision Support Systems*, *86*, 35–47.

Wang, R., Edgar, T. F., Baldea, M., Nixon, M., Wojsznis, W., & Dunia, R. (2017). A geometric method for batch data visualization, process monitoring and fault detection. *Journal of Process Control*, *67*, 197–205.

Wang, R. Y. (1998). A Product Perspective on Total Data Quality Management. *Commun. ACM*, *41*(2), 58–65.

Wang, R. Y., Storey, V. C., & Firth, C. P. (1995). A Framework for Analysis of Data Quality Research. *IEEE Transaction on Knowledge and Data Engineering*, *7*(4).

Wang, R. Y., & Strong, D. M. (1996). Beyond Accuracy: What Data Quality Means to Data Consumers. *J. Manage. Inf. Syst.*, *12*(4), 5–33. Retrieved from http://dl.acm.org/citation.cfm?id=1189570.1189572

Wang, W., Liu, L., Feng, Y., & Wang, T. (2014). Innovation with IS usage: individual absorptive capacity as a mediator. *Industrial Management & Data Systems*, *144*(8), 1110–1130.

Wang, Y., Kung, L., & Byrd, T. A. (2016). Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations. *Technological Forecasting and Social Change*.

Warth, J., Kaiser, G., & Kügler, M. (2011). The impact of data quality and analytical capabilities on planning performance: insights from the automotive industry. In *Tenth International Conference on Wirtschaftsinformatik. Association for Information Systems Electronic Library* (pp. 322–331). Zurich, Switzerland. Retrieved from http://www.wi2011.ch

Watson, H. J., & Wixom, B. H. (2007). The current state of business intelligence. *IEEE Computer Society*, *40*(9), 96–99.

Watts, C., Hahn, C., & Sohn, B. (1993). Manufacturing flexibility: concept and measurement. *Operations Management Review*, *9*(4), 33–44.

Wei, J., Lowry, P. B., & Seedorf, S. (2015). The assimilation of RFID technology by Chinese companies: A technology diffusion perspective. *Information & Management*, *52*(6), 628–642.

Wei, Z., Song, X., & Wang, D. (2017). Manufacturing flexibility, business model design, and firm performance. *International Journal of Production Economics*, *193*(28), 87–97.

Wheelen, T. L., Hunger, J. D., Hoffman, A. N., & Bamford, C. E. (2018). *Strategic Management and Business Policy : Globalization, Innovation and Sustainability* (15th ed.). Pearson Education Limited.

White, A., Daniel, E. M., & Mohdzain, M. (2005). The role of emergent information technologies and systems in enabling supply chain agility. *International Journal of Information Management*, *25*(5), 396–410.

Whitehead, K. K., Zacharia, Z. G., & Prater, E. L. (2016). Absorptive capacity versus distributive capability: The asymmetry of knowledge transfer. *International Journal of Operations & Production Management*, *36*(10), 1308–1332.

Wiengarten, F., Humphreys, P., Cao, G., Fynes, B., & McKittrick, A. (2010). Collaborative supply chain practices and performance: exploring the key role of information quality. *Supply Chain Management: An International Journal*, *15*(6), 463–473.

Winter, S. G. (2003). Understanding dynamic capabilities. *Strategic Management Journal*, *24*(10 SPEC ISS.), 991–995.

Wolfe, R. A. (1994). Organizational Innovation: Review, Critique and Suggested Research Directions. *Journal of Management Studies*, *31*(3), 405–431.

Woodall, P., Borek, A., & Parlikad, A. K. (2013). Data quality assessment: The Hybrid Approach. *Information and Management*, *50*(7), 369–382.

Wu, I.-L., & Chuang, C.-H. (2010). Examining the diffusion of electronic supply chain management with external antecedents and firm performance: A multi-stage analysis. *Decision Support Systems*, *50*(1), 103–115.

Wu, K.-L., & Yang, M.-S. (2005). A cluster validity index for fuzzy clustering. *Pattern Recognition Letters*, *26*(9), 2176–2177.

Wu, S. J., Melnyk, S. A., & Flynn, B. B. (2010). Operational Capabilities: The Secret

Ingredient. *Decision Sciences*, *41*(4), 721–754.

Wu, T., Jim Wu, Y.-C., Chen, Y. J., & Goh, M. (2014). Aligning supply chain strategy with corporate environmental strategy: A contingency approach. *International Journal of Production Economics*, *147*, 220–229.

Xie, X. L., & Beni, G. (1991). A validity measure for fuzzy clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Xiong, G., Nyberg, T. R., Dong, X., & Shang, X. (2015). Intelligent technologies and systems of material management. In C. Kahraman & S. Çevik Onar (Eds.), *Intelligent Systems Reference Library* (Vol. 87, pp. 295–330). Cham: Springer International Publishing.

Xu, R., & Wunsch, D. (2005). Survey of clustering algorithms. *IEEE Transactions on Neural Networks*, *16*(3), 645–678.

Xu, Z., Frankwick, G. L., & Ramirez, E. (2015). Effects of big data analytics and traditional marketing analytics on new product success: A knowledge fusion perspective. *Journal of Business Research*.

Yang, Y., Lee, P. K. C., & Cheng, T. C. E. (2017). Leveraging selected operational improvement practices to achieve both efficiency and creativity: A multi-level study in frontline service operations. *International Journal of Production Economics*, *191*, 298–310.

Yao, Y., Dresner, M., & Palmer, J. W. (2009). Impact of boundary-spanning information technology and position in chain on firm performance. *Journal of Supply Chain Management*, *45*(4), 3–17.

Yesudas, M., Menon, G., & Ramamurthy, V. (2014). Intelligent operational dashboards for smarter commerce using big data. *IBM Journal of Research and Development*, *58*(5/6), 13:1-13:10.

Yeung, A. C. L., Cheng, T. C. E., & Kee-hung, L. (2005). An Empirical Model for Managing Quality in the Electronics Industry. *Production and Operations Management*, *14*(2), 189–204.

Yin, R. K. (2003). Case Study Research Design and Methods. *SAGE Publications*.

Yin, R. K. (2008). *Case Study Research: Design and Methods*. *Sage Publications* (4th ed.). Sage Publications, Inc.

Yong, A. G., & Pearce, S. (2013). A Beginner’s Guide to Factor Analysis: Focusing on Exploratory Factor Analysis. *Tutorials in Quantitative Methods for Psychology*, *9*(2), 79–94.

Youndt, M. A., Snell, S. A., Dean, J. W., & Lepak, D. P. (1996). Human resource management, manufacturing strategy, and firm performance. *Academy of Management Journal*, *39*(4), 836–865.

Yu, W. (2015). The effect of IT-enabled supply chain integration on performance. *Production Planning & Control*, *26*(12), 945–957.

Yu, W., Chavez, R., Jacobs, M. A., & Feng, M. (2017). Data-driven supply chain capabilities and performance : A resource-based view. *Transportation Research Part*

Yu, W., Jacobs, M. A., Salisbury, W. D., & Enns, H. (2013). The effects of supply chain integration on customer satisfaction and financial performance: An organizational learning perspective. *International Journal of Production Economics*, *146*(1), 346–358.

Zacharia, Z. G., Nix, N. W., & Lusch, R. F. (2011). Capabilities that enhance outcomes of an episodic supply chain collaboration. *Journal of Operations Management*, *29*(6), 591–603.

Zahra, S. A., & George, G. (2002). Absorptive Capacity : A Review , Reconceptualization , and Extension. *The Academy of Management Review*, *27*(2), 185–203.

Zahra, S., & George, G. (2002a). Absorptive Capacity : A Review , Reconceptualization. *The Academy of Management Review*, *27*(2), 185–203.

Zahra, S., & George, G. (2002b). The Net-enabled business innovation cycle and a strategic entrepreneurship perspective on the evolution of dynamic capabilities. *Information Systems Research*, *13*(2), 147–150.

Zailani, S., Govindan, K., Iranmanesh, M., & Shaharudin, M. R. (2015). Green Innovation Adoption in Automotive Supply Chain: The Malaysian case. *Journal of Cleaner Production*, *108*, 1115–1122.

Zhang, M., Zhao, X., Lyles, M. A., & Guo, H. (2015). Absorptive capacity and mass customization capability. *International Journal of Operations & Production Management*, *35*(9), 1275–1294.

Zhang, Q., Huang, T., Zhu, Y., & Qiu, M. (2013). A Case Study of Sensor Data Collection and Analysis in Smart City: Provenance in Smart Food Supply Chain. *International Journal of Distributed Sensor Networks*, 1–12.

Zhang, X., Donk, D. P. Van, & Vaart, T. van der. (2016). The different impact of inter-organizational and intra-organizational ICT on supply chain performance. *International Journal of Operations & Production Management*, *36*(7), 803–824.

Zhang, Y., Ren, S., Liu, Y., & Si, S. (2017). A big data analytics architecture for cleaner manufacturing and maintenance processes of complex products. *Journal of Cleaner Production*, *142*, 626–641.

Zhao, X., Lynch, J. G., & Chen, Q. (2010). Reconsidering Baron and Kenny: Myths and Truths about Mediation Analysis. *Journal of Consumer Research*, *37*(2), 197–206.

Zheng, L., Tang, L., Li, T., Duan, B., Lei, M., Wang, P., … Li, H. (2014). Applying data mining techniques to address critical process optimization needs in advanced manufacturing. *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '14*, 1739–1748.

Zhong, R. Y., Huang, G. Q., Lan, S., Dai, Q. Y., Chen, X., & Zhang, T. (2015). A Big Data Approach For Logistics Trajectory Discovery From Rfid-Enabled Production Data. *International Journal of Production Economics*, *165*, 260–272.

Zhong, R. Y., Lan, S., Xu, C., Dai, Q., & Huang, G. Q. (2016). Visualization of RFID-enabled shopfloor logistics Big Data in Cloud Manufacturing. *International Journal of Advanced Manufacturing Technology*, *84*(1), 5–16.

Zhong, R. Y., Newman, S. T., Huang, G. Q., & Lan, S. (2016). Big Data for supply chain management in the service and manufacturing sectors: Challenges, opportunities, and future perspectives. *Computers & Industrial Engineering*.

Zikmund, W. G., Babin, B. J., Carr, J. C., & Griffin, M. (2010). *Business Research Methods* (8th ed.).

Zou, X., Tao, F., Jiang, P., Gu, S., Qiao, K., Zuo, Y., & Xu, L. (2016). A new approach for data processing in supply chain network based on FPGA. *The International Journal of Advanced Manufacturing Technology*, *84*(4), 249–260.

Appendices

Appendix A: Questionnaire

A survey on the impact of Big Data and Business Analytics on Firm performance

Dear Sir/Madam

The University of Sheffield Management School is conducting this study to understand the impact of **Big Data and Business Analytics (BDA)** on **firm performance**. This research aims to enhance the understanding of BDA practices and its potential to create business value. This study has received ethical approval from the University of Sheffield. The questionnaire will take about **20 minutes to complete**. Your responses will be treated in the **strictest confidence** and your personal information will **not be disclosed to anyone**. Your participation in this study is **voluntary** and you can withdraw from the study at any time simply by closing your browser. A **summary report of findings** will be sent to all those involved in this study. This report will be **beneficial in maintaining and enhancing your long-term competitiveness** As a small token of appreciation, every respondent will be entered in a draw to win an **Amazon gift voucher worth £100**. Please do not forget to **include your email address** at the end of your responses to receive a summary report of this study and to enter into the draw.

Thank you for your time and support.

Yours Sincerely

Mr. Deepak Arunachalam

Email: darunachalam1@sheffield.ac.uk

Phone: +44(0) 7459850512

Research Supervisors: Dr. Andrew Brint (a.brint@sheffied.ac.uk),

Dr. Chantal C. Cantarelli (c.c.cantarelli@sheffield.ac.uk)

Dr. Niraj Kumar(niraj.kumar@liverpool.ac.uk

Big Data Analytics: What does it mean?

Before we can start, it is important to explain what is Big Data Analytics. Every time you base a decision on facts and data, analytics is involved. Every time you receive a report with nice graphs and colors via interactive data visualisation, analytics is involved. Different terms such as Business intelligence/Business analytics are used in the industry to represent data and analytics practices. But, due to the increasing complexity of analysing large, complex data, the concept of 'Big Data' and advanced analytics has emerged as an innovative technology and widely used by organisations.

"**Big Data**" refers to large structured and unstructured data sets that require new forms of processing capability to enable better decision making. Some of the examples could include sales data, process operating data and other information captured by sensors, web server logs, Internet clickstream, social media activity, mobile-phone call records, etc. Whereas, "**Big Data Analytics**" is the process of examining big data using analytics tools and techniques (such as regression, time series analysis, clustering, support vector machines, neural networks), which is seen by organisations as a significant tool to improve operational efficiency, develop new revenue streams and gain sustainable competitive advantage

Section A: Background Information

Please indicate your answer to each question by either writing in the space provided or ticking the most appropriate option.

Q1. What is the **primary industry** in which your organisation operates?

- ○ Information and communication
- ○ Manufacturing
- ○ Retail/Wholesale
- ○ Transportation and storage
- ○ Water supply, sewerage, waste management
- ○ Others (Please specify) _______________

Q2. Which **manufacturing sector** does your organisation belong to?

- ○ Automobiles
- ○ Textiles
- ○ Pharmaceuticals
- ○ Rubber and tube industries
- ○ Metals
- ○ Chemicals
- ○ Computer and electronic products
- ○ Papers and leathers
- ○ Food and dairy products
- ○ Beverages
- ○ Electrical equipment
- ○ Others (please specify) ____________

Q3 Please indicate your level of awareness about your organisation's data and analytics resources, decision-making processes and supply chain activities: (1-Not aware at All to 5-Extremely Aware)	1	2	3	4	5
Your awareness about the organisation's ability to use data and analytics resources?	○	○	○	○	○
Your awareness about the organisation’s decision-making processes?	○	○	○	○	○
Your awareness about the organisation’s supply chain related activities?		○	○	○	○

Section B: Your Big Data Analytics Capabilities

In this section, please focus on your organisation's capabilities to use data and analytics tools/ techniques for decision-making.

Q4. To what extent do you disagree or agree with the following statements regarding your organisation's ability to generate/access/collect data? (1-strongly disagree to 5-strongly agree)	1	2	3	4	5
We are able to generate/access/collect a large volume of data (from our operational systems and other data sources).	○	○	○	○	○
We are able to generate/access/collect unstructured data -different kinds of data (such as textual data, audio and visual data, images, sensor data, RFID data and social media data).	○	○	○	○	○
We are able to generate/access/collect data in real-time.	○	○	○	○	○
We are able to generate/access/collect data from a wide range of data sources (heterogeneity).	○	○	○	○	○
We have access to very large, unstructured, or fast-moving data for analysis.	○	○	○	○	○

Q5. To what extent do you disagree or agree with the following statements regarding your organisation's ability to integrate and store data? (1-strongly disagree to 5-strongly agree)	1	2	3	4	5
We integrate data from multiple internal sources into a data warehouse or data mart for easy access.	○	○	○	○	○
We integrate external data with internal data to facilitate high-value analysis of our business environment. We have ability to Extract, Transform and Load (ETL) data from across systems and organisational boundaries.	○	○	○	○	○
In our organisation, data is not integrated or poorly integrated	○	○	○	○	○
Our data storage system is able to manage large volume of data.	○	○	○	○	○
Our data storage system is able to manage different data types beyond structured data.	○	○	○	○	○
We have adopted parallel computing approaches (e.g., Hadoop) for big data processing.	○	○	○	○	○
We have adopted new forms of databases such as Not Only SQL (NoSQL) for storing and retrieving of data.	○	○	○	○	○

Q6. To what extent do you **disagree or agree** with the following statements regarding your organisation's ability to **use analytics tools and techniques for decision-making at intra-organisational level**? (1-strongly disagree to 5-strongly agree)	1	2	3	4	5
We use analytics to get data-driven insights into our historical business performance.	○	○	○	○	○
We use analytics to predict future events of our business environment.	○	○	○	○	○
We use analytics to prescribe possible courses of action for our business.	○	○	○	○	○
The performance of analytic models is regularly reviewed once deployed.	○	○	○	○	○
We are able to use analytical tools to convert data into actionable insights.	○	○	○	○	○

Q7: How often does your organization use the following analytics techniques? (1- Never to 5- Always)	1	2	3	4	5
Web analytics	○	○	○	○	○
Social media analytics	○	○	○	○	○
Text, audio, video analytics	○	○	○	○	○
Data and text mining	○	○	○	○	○

Q8. To what extent has your organisation use Big Data Analytics to manage the following supply chain functions? (1- Little or no usage to 5- Heavy usage)	1	2	3	4	5
Sourcing analysis	○	○	○	○	○
Purchasing spend analytics	○	○	○	○	○
CRM/customer analysis	○	○	○	○	○
Network design/optimisation	○	○	○	○	○
Warehouse operations improvements	○	○	○	○	○
Process monitoring	○	○	○	○	○
Production Process optimization	○	○	○	○	○
Logistics process improvements	○	○	○	○	○
Forecasting/demand management	○	○	○	○	○
Inventory optimization	○	○	○	○	○

Q9. To what extent do you **disagree or agree** with the following statements regarding your organisation's **ability to visualize data and information effectively**? (1-strongly disagree to 5-strongly agree)	1	2	3	4	5
We have adopted data visualization tools.	○	○	○	○	○
Data-driven insights are delivered using dashboards or other interactive visualisation tools.	○	○	○	○	○
Data-driven insights are delivered in such a way that they are easily understandable by the target group.	○	○	○	○	○
Data-driven insights are delivered in real-time.	○	○	○	○	○
Data-driven insights are shared seamlessly across our organisation, regardless of the location.	○	○	○	○	○

Q10. To what extent do you disagree or agree with the following statements regarding perception and attitude towards the use of Big Data Analytics and data-driven decision-making in your organisation? (1-strongly disagree to 5-strongly agree)	1	2	3	4	5
We consider data as a tangible asset.	○	○	○	○	○
We are willing to override our own intuition when data contradicts our viewpoints.	○	○	○	○	○
We base our decisions on data rather than on instinct.	○	○	○	○	○
We continuously assess and improve our practices in response to insights extracted from data.	○	○	○	○	○
We continuously coach our employees to make decisions based on data.	○	○	○	○	○

Q11. To what extent do you disagree or agree with the following statements regarding your organisation's ability to possess Big Data Analytics skills? (1-strongly disagree to 5-strongly agree)	1	2	3	4	5
We provide big data analytics training to our own employees.	○	○	○	○	○
We hire new employees that already have big data analytics skills.	○	○	○	○	○
Our big data analytics staff have the right skills to accomplish their jobs successfully.	○	○	○	○	○
Our big data analytics staff have suitable education to fulfill their jobs.	○	○	○	○	○
Our big data analytics staff holds suitable work experience to accomplish their jobs successfully.	○	○	○	○	○
Our big data analytics staff are well trained.	○	○	○	○	○

Q12. Please rate your organisation's ability to maintain to the quality of data products/information on the following aspects (1-Poor to 5-Excellent)	1	2	3	4	5
Completeness	○	○	○	○	○
Timeliness	○	○	○	○	○
Reliability	○	○	○	○	○
Consistency	○	○	○	○	○
Accuracy	○	○	○	○	○

Section C: Your organisation's learning ability

Q13. To what extent do you disagree or agree with the following statements regarding your organisations ability to acquire, integrate and exploit knowledge? (1-strongly disagree to 5-strongly agree)	1	2	3	4	5
We are successful in learning new things.	○	○	○	○	○
We are effective in developing new knowledge or insights that have the potential to influence product/service development.	○	○	○	○	○
We are able to identify and acquire internal (e.g., within the firm) and external (e.g., market) knowledge.	○	○	○	○	○
We have effective routines to identify, value, and import new information and knowledge from channel partners.	○	○	○	○	○
We have adequate routines to analyse the information and knowledge obtained.	○	○	○	○	○
We have adequate routines to assimilate new information and knowledge.	○	○	○	○	○
We can successfully integrate our existing knowledge with the new information and knowledge acquired.	○	○	○	○	○
We are effective in transforming existing information into new knowledge.	○	○	○	○	○
We can successfully grasp the opportunities for our firm from new external knowledge.	○	○	○	○	○
We can successfully exploit the new integrated information and knowledge into concrete applications.	○	○	○	○	○
We are effective in utilising knowledge into new products.	○	○	○	○	○
We constantly consider better ways to exploit knowledge.	○	○	○	○	○

Section D: Your Organisation's Performance

In this section, please focus on your organisation's operational and innovation performance.

Q14: Please indicate your opinion about how well your organisation performs compared with industry benchmark (1-Poor to 5-Excellent)	1	2	3	4	5
Quality-related performance					
Product quality and performance	○	○	○	○	○
Conformance to product specifications (meet established standards/customers' requirements)	○	○	○	○	○
Products reliability (probability of a product malfunctioning/failing within a specified time period)	○	○	○	○	○
Cost-related performance					
Unit cost of manufacturing	○	○	○	○	○
Inventory costs	○	○	○	○	○
Overhead costs	○	○	○	○	○
Price competitiveness	○	○	○	○	○
Flexibility-related performance					
Flexibility to change product mix	○	○	○	○	○
Flexibility to change volume	○	○	○	○	○
Ability to produce customized product features	○	○	○	○	○
Time-based performance					
On-time delivery performance	○	○	○	○	○
Accuracy of product delivery (correct quantity and product)	○	○	○	○	○
Fast/quick delivery of products	○	○	○	○	○
Order fulfillment lead-time	○	○	○	○	○
Manufacturing lead-time	○	○	○	○	○
Supply chain throughput time	○	○	○	○	○

Q15. To what extent do you disagree or agree with the following statements regarding your organisation's innovation performance compared with industry benchmark? (1-strongly disagree to 5-strongly agree)	1	2	3	4	5
We achieve substantial innovations in product and/or service.	○	○	○	○	○
We achieve substantial innovations in management.	○	○	○	○	○
We achieve substantial innovations in manufacturing technology.	○	○	○	○	○

Section E: Your Demographic information

Q16. E1. How **many employees work** in your establishment?

○ 1-9 employees

○ 10-49 employees

○ 50-249 employees

○ 250-500 employees

○ More than 500 employees

Q17. Please indicate the **annual turnover** of your organisation (in British Pound £):

○ ≤ 2 Million

○ 2–10 million

○ 10–25 million

○ 25 - 50 million

○ > 50 million

Q18. What is **your job-title in** the organisation?

- ◯ Chief Executive officer
- ◯ Chief Information officer
- ◯ General Manager
- ◯ Director
- ◯ Vice President
- ◯ Senior Manager (Product Development/ Operations/ Supply chain/ Logistics /Waste management etc.)
- ◯ Senior Manager (Information Technology/Big Data Analytics/supply chain analytics)
- ◯ Manager (Product Development/ Operations/ Supply chain/ Logistics / Waste management etc.)
- ◯ Manager (Information Technology/Big Data Analytics/supply chain analytics)
- ◯ Others (please specify) ____________________

Q19. How **long have you been working** in this organisation?

- ◯ Less than 6 months
- ◯ 6 - 12 months
- ◯ 1 - 5 years
- ◯ 5 - 10 years
- ◯ 10 -15 years
- ◯ 15-20 years
- ◯ More than 20 years

Q20. Please select your **organisation's location** in the United Kingdom **(select all that apply)**:

- ○ North East England
- ○ North West England
- ○ Yorkshire & Humber
- ○ East Midlands
- ○ West Midlands
- ○ East of England
- ○ London
- ○ South East England
- ○ South West England
- ○ Wales
- ○ Scotland
- ○ Northern Ireland

Section F: Final Section

Q21. Please enter your email address, if you would like to receive a summary of the study findings and to include in a prize draw:

Appendix B: Ethics approval letter

The University Of Sheffield.

Downloaded: 08/08/2016
Approved: 08/08/2016

Deepak Arunachalam
Registration number: 150138156
Management School
Programme: PhD

Dear Deepak

PROJECT TITLE: Big Data Analytics practice in the UK manufacturing industry
APPLICATION: Reference Number 008148

On behalf of the University ethics reviewers who reviewed your project, I am pleased to inform you that on 08/08/2016 the above-named project was **approved** on ethics grounds, on the basis that you will adhere to the following documentation that you submitted for ethics review:

- University research ethics application form 008148 (dated 26/07/2016).
- Participant information sheet 1022437 version 1 (26/07/2016).
- Participant information sheet 1022436 version 1 (26/07/2016).
- Participant information sheet 1017052 version 1 (11/04/2016).
- Participant consent form 1022439 version 1 (26/07/2016).
- Participant consent form 1022438 version 1 (26/07/2016).
- Participant consent form 1017048 version 2 (26/07/2016).

If during the course of the project you need to deviate significantly from the above-approved documentation please inform me since written approval will be required.

Yours sincerely

Daniel Miller
Ethics Administrator
Management School

Appendix C: Pre-test questions / Feedback

Please provide your feedback on the questionnaire content and design.

Q1. How long do you take to fill the complete form (in minutes)?

Q2. Are the instructions clear enough to answer the questions asked in the survey?

- ❍ Yes
- ❍ No (Please specify in the box below) ____________________

Q3. Are the questions clear and easily understandable?

- ❍ Yes
- ❍ No (Please specify in the box below) ____________________

Q4. Are there any problems in understanding what kind of answers are expected, or in providing answers to the questions posed?

- ❍ Yes
- ❍ No (Please specify in the box below) ____________________

Q5. Are the structure (i.e: sequence of questions) and design (i.e: font size, the spacing between sentences etc.) of the questionnaire logical?

- ❍ Yes
- ❍ No (Please specify in the box below) ____________________

Q6. Is the questionnaire easy to complete?

- ❍ Yes
- ❍ No (Please specify in the box below) ____________________

Q7. Is the language used in the questionnaire free from jargon, slang, and abbreviation?

- ❍ Yes
- ❍ No (Please specify in the box below) ____________________

Q8. Is there any question, which the respondent objects to answer?

- ❍ Yes (Please specify in the box below) ____________________
- ❍ No

Q9. Is the length of the questionnaire appropriate?

- ❍ Yes
- ❍ No (Please specify in the box below) ____________________

Q10. Are there any elements of the questionnaire, which you think should be changed, deleted or modified?

- ❍ Yes (Please specify in the box below) ____________________
- ❍ No

Finally, please provide your overall reaction to the questionnaire (i.e: is there anything you don't like about the questionnaire, or any suggestion to improve the questionnaire and response rate).

Appendix D: Data background

Table 7.1 Pre-screening question 1

SQ1					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Somewhat aware	57	25.8	25.8	25.8
	Moderately aware	64	29.0	29.0	54.8
	Extremely aware	100	45.2	45.2	100.0
	Total	221	100.0	100.0	

Table 7.2 Pre-screening question 2

SQ2					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Somewhat aware	36	16.3	16.3	16.3
	Moderately aware	52	23.5	23.5	39.8
	Extremely aware	133	60.2	60.2	100.0
	Total	221	100.0	100.0	

Table 7.3 Pre-screening question 3

SQ3					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Somewhat aware	48	21.7	21.7	21.7
	Moderately aware	81	36.7	36.7	58.4
	Extremely aware	92	41.6	41.6	100.0
	Total	221	100.0	100.0	

Appendix E Outputs of Factor analysis

Table 7.4 Communalities of the list of items measured

Communalities		
	Initial	Extraction
DG2	0.762	0.687
DG4	0.820	0.827
DG5	0.798	0.821
DIM1	0.776	0.718
DIM2	0.802	0.751
DIM3	0.788	0.762
DIM4RC	0.739	0.622
DIM6	0.702	0.610
DIM7	0.773	0.618
DIM8	0.734	0.576
AA1	0.833	0.762
AA2	0.854	0.846
AA3	0.792	0.794
AA4	0.783	0.680
AA5	0.825	0.774
WEBAnalytics	0.797	0.753
SMAnaly	0.794	0.740
TAVanaly	0.778	0.746
SCA1	0.750	0.671
SCA2	0.790	0.661
SCA3	0.724	0.616
SCA4	0.677	0.556
SCA5	0.765	0.645
SCA6	0.798	0.709
SCA7	0.801	0.760
SCA8	0.768	0.698
SCA9	0.800	0.735
SCA10	0.779	0.679
DV1	0.890	0.876
DV2	0.874	0.873
DV3	0.866	0.839
DV5	0.873	0.820
DDC1	0.743	0.690
DDC2	0.773	0.724
DDC3	0.768	0.714
DDC4	0.813	0.781
DDC5	0.773	0.740
BDS3	0.901	0.872

BDS4	0.891	0.876
BDS5	0.899	0.876
BDS6	0.882	0.855
DIQ1	0.767	0.639
DIQ2	0.771	0.646
DIQ3	0.858	0.821
DIQ4	0.849	0.821
DIQ5	0.830	0.814
ACAP1	0.867	0.763
ACAP2	0.873	0.763
ACAP3	0.880	0.773
ACAP4	0.798	0.730
ACAP5	0.872	0.754
ACAP6	0.883	0.796
ACAP7	0.883	0.813
ACAP8	0.879	0.815
ACAP9	0.879	0.783
ACAP10	0.877	0.778
ACAP11	0.860	0.804
ACAP12	0.844	0.786
Extraction Method: Maximum Likelihood.		

Table 7.5 EFA of organisational capabilities dimensions -Total variance explained

Total Variance Explained							
Factor	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings[a]
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total
1	26.281	45.311	45.311	25.773	44.437	44.437	21.240
2	3.549	6.118	51.430	3.043	5.246	49.682	19.014
3	2.450	4.225	55.654	1.869	3.222	52.905	15.651
4	2.342	4.038	59.692	1.765	3.043	55.948	14.561
5	1.943	3.350	63.042	1.969	3.394	59.342	9.292
6	1.838	3.169	66.211	1.690	2.914	62.256	13.926
7	1.494	2.576	68.787	1.522	2.624	64.881	18.101
8	1.281	2.208	70.995	1.120	1.931	66.812	13.661
9	1.233	2.126	73.121	1.103	1.902	68.714	11.021
10	1.117	1.926	75.047	1.008	1.737	70.451	10.550
11	0.956	1.648	76.695				
12	0.833	1.437	78.132				
13	0.721	1.243	79.375				
14	0.685	1.180	80.555				
15	0.621	1.071	81.626				
16	0.597	1.029	82.655				
17	0.543	0.936	83.591				
18	0.508	0.877	84.468				
19	0.493	0.850	85.318				
20	0.470	0.811	86.128				
21	0.441	0.761	86.889				
22	0.427	0.736	87.625				
23	0.407	0.701	88.327				
24	0.374	0.645	88.972				
25	0.370	0.639	89.610				
26	0.343	0.591	90.202				
27	0.332	0.573	90.775				
28	0.317	0.547	91.322				
29	0.291	0.502	91.823				
30	0.286	0.493	92.316				
31	0.275	0.475	92.791				
32	0.257	0.443	93.233				
33	0.246	0.423	93.657				
34	0.237	0.408	94.065				
35	0.223	0.384	94.449				
36	0.218	0.376	94.825				
37	0.211	0.363	95.188				
38	0.205	0.353	95.542				

39	0.201	0.347	95.889				
40	0.196	0.338	96.227				
41	0.193	0.333	96.560				
42	0.179	0.309	96.869				
43	0.168	0.290	97.159				
44	0.159	0.273	97.433				
45	0.149	0.256	97.689				
46	0.144	0.248	97.937				
47	0.136	0.234	98.171				
48	0.130	0.224	98.395				
49	0.127	0.219	98.614				
50	0.110	0.189	98.803				
51	0.105	0.181	98.984				
52	0.102	0.176	99.159				
53	0.097	0.168	99.327				
54	0.094	0.162	99.489				
55	0.087	0.150	99.639				
56	0.077	0.133	99.771				
57	0.071	0.122	99.893				
58	0.062	0.107	100.000				
Extraction Method: Maximum Likelihood.							

Table 7.6 EFA of firm performance dimensions -Total variance explained

Total Variance Explained							
Factor	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings[a]
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total
1	4.944	29.079	29.079	1.973	11.605	11.605	4.364
2	4.156	24.448	53.527	4.710	27.708	39.313	3.331
3	1.678	9.868	63.396	2.635	15.500	54.813	3.220
4	1.204	7.084	70.480	1.454	8.551	63.363	2.892
5	1.030	6.062	76.542	0.795	4.677	68.040	1.931
6	0.781	4.596	81.137				
7	0.531	3.123	84.260				
8	0.428	2.518	86.778				
9	0.389	2.288	89.067				
10	0.344	2.023	91.089				
11	0.324	1.908	92.997				
12	0.261	1.534	94.531				
13	0.223	1.313	95.844				
14	0.209	1.228	97.072				
15	0.176	1.038	98.110				
16	0.171	1.004	99.114				

17	0.151	0.886	100.000
Extraction Method: Maximum Likelihood.			

Table 7.7 Common latent factor bias test results

Standardized Regression Weights: (Default model)			**Standardized Regression Weights: (Common Latent factor Model- Default model)**			
Items	Factors	Estimate	Items	Factors	Estimate	Delta
OP12	Time	0.921	OP12	Time	0.817	0.104
OP11	Time	0.873	OP11	Time	0.756	0.117
OP13	Time	0.901	OP13	Time	0.748	0.153
DDC5	Data culture	0.855	DDC5	Data culture	0.713	0.142
DDC2	Data culture	0.831	DDC2	Data culture	0.733	0.098
DDC4	Data culture	0.879	DDC4	Data culture	0.707	0.172
BDS5	BDataskills	0.928	BDS5	BDataskills	0.774	0.154
BDS4	BDataskills	0.927	BDS4	BDataskills	0.763	0.164
BDS6	BDataskills	0.917	BDS6	BDataskills	0.726	0.191
BDS3	BDataskills	0.934	BDS3	BDataskills	0.766	0.168
DG5ALL	Dgenera	0.878	DG5ALL	Dgenera	0.77	0.108
DG4VSources	Dgenera	0.914	DG4VSources	Dgenera	0.736	0.178
DG2Vareity	Dgenera	0.827	DG2Vareity	Dgenera	0.646	0.181
OP6	costperf	0.824	OP6	costperf	0.718	0.106
OP5	costperf	0.866	OP5	costperf	0.707	0.159
OP4	costperf	0.815	OP4	costperf	0.658	0.157
OP1	ProdQual	0.871	OP1	ProdQual	0.692	0.179
OP2	ProdQual	0.9	OP2	ProdQual	0.709	0.191
OP3	ProdQual	0.928	OP3	ProdQual	0.769	0.159
DV2	DataViz	0.924	DV2	DataViz	0.792	0.132
DV3	DataViz	0.901	DV3	DataViz	0.735	0.166
DV1	DataViz	0.93	DV1	DataViz	0.783	0.147
AA3	AdvAnal	0.829	AA3	AdvAnal	0.718	0.111
AA2	AdvAnal	0.922	AA2	AdvAnal	0.775	0.147
AA1	AdvAnal	0.869	AA1	AdvAnal	0.732	0.137
INNOV1	innovperfo	0.904	INNOV1	innovperfo	0.738	0.166
INNOV2	innovperfo	0.886	INNOV2	innovperfo	0.706	0.18
INNOV3	innovperfo	0.841	INNOV3	innovperfo	0.685	0.156
DIM8	DIMCAp	0.695	DIM8	DIMCAp	0.59	0.105
DIM3	DIMCAp	0.857	DIM3	DIMCAp	0.756	0.101
DIM7	DIMCAp	0.698	DIM7	DIMCAp	0.56	0.138
OP8	Flexibilityperfo	0.885	OP8	Flexibilityperfo	0.793	0.092
OP10	Flexibilityperfo	0.793	OP10	Flexibilityperfo	0.714	0.079

OP9	Flexibilityperfo	0.855	OP9	Flexibilityperfo	0.719	0.136
SCA2	SCACAP	0.772	SCA2	SCACAP	0.655	0.117
SCA5	SCACAP	0.799	SCA5	SCACAP	0.581	**0.218**
SCA8	SCACAP	0.836	SCA8	SCACAP	0.721	0.115
SCA9	SCACAP	0.823	SCA9	SCACAP	0.69	0.133
SCA4	SCACAP	0.701	SCA4	SCACAP	0.528	0.173
SCA1	SCACAP	0.761	SCA1	SCACAP	0.576	0.185
ACAP1	ACAP_fac	0.841	ACAP1	ACAP_fac	0.681	0.16
ACAP2	ACAP_fac	0.839	ACAP2	ACAP_fac	0.572	**0.267**
ACAP3	ACAP_fac	0.842	ACAP3	ACAP_fac	0.653	0.189
ACAP4	ACAP_fac	0.834	ACAP4	ACAP_fac	0.65	0.184
ACAP5	ACAP_fac	0.834	ACAP5	ACAP_fac	0.73	0.104
ACAP6	ACAP_fac	0.861	ACAP6	ACAP_fac	0.692	0.169
ACAP7	ACAP_fac	0.906	ACAP7	ACAP_fac	0.725	0.181
ACAP8	ACAP_fac	0.9	ACAP8	ACAP_fac	0.84	0.06
ACAP9	ACAP_fac	0.869	ACAP9	ACAP_fac	0.761	0.108
ACAP10	ACAP_fac	0.871	ACAP10	ACAP_fac	0.772	0.099
ACAP11	ACAP_fac	0.878	ACAP11	ACAP_fac	0.69	0.188
ACAP 12	ACAP_fac	0.875	ACAP 12	ACAP_fac	0.71	0.165
DIQ1	DQual	0.821	DIQ1	DQual	0.719	0.102
DIQ2	DQual	0.821	DIQ2	DQual	0.672	0.149
DIQ3	DQual	0.9	DIQ3	DQual	0.745	0.155
DIQ4	DQual	0.917	DIQ4	DQual	0.764	0.153
Textaudiovideoanalytics	DigAnal	0.871	Textaudiovideoanalytics	DigAnal	0.718	0.153
Socialmediaanalytics	DigAnal	0.821	Socialmediaanalytics	DigAnal	0.711	0.11
Webanalytics	DigAnal	0.796	Webanalytics	DigAnal	0.66	0.136
DIM1	DIMCAp	0.852	DIM1	DIMCAp	0.756	0.096
Dataandtextmining	DigAnal	0.866	Dataandtextmining	DigAnal	0.716	0.15
DIM4RCode	DIMCAp	0.773	DIM4RCode	DIMCAp	0.632	0.141
DIQ5	DQual	0.882	DIQ5	DQual	0.715	0.167
DDC3	Dataculture	0.822	DDC3	Dataculture	0.701	0.121
DDC1	Dataculture	0.809	DDC1	Dataculture	0.685	0.124
DV5	DataViz	0.9	DV5	DataViz	0.706	0.194
DIM2	DIMCAp	0.845	DIM2	DIMCAp	0.712	0.133
OP14	Time	0.827	OP14	Time	0.658	0.169
OP7	costperf	0.789	OP7	costperf	0.604	0.185

Appendix F: Systematic literature review of BDA maturity models

Reference	Origin	No of Dimensions	Dimensions	Stages	Stages	Assessment method	Theoretical relevance	Model Design method	Focus of Maturity
IDC (2013)	Business	5	Intent(Strategy, Sponsorship, justification), Data (Relevance, Quality, Availability), Technology(Functionality, Performance, Adoption), Process(Tracking, Analysis, Decision), People(Skill, Culture and Organisation structure)	5	Ad-Hoc, Opportunistic, Repeatable, Managed, Optimised	Questionnaire	n/a	n/a	Process, Object, People
Nott (2014)	Business	6	Business strategy, Information, Analytics, Culture and execution, Architecture, Governance	5	Ad-Hoc, Foundational, Competitive, Differentiating, Breakaway	Textual-Descriptive	n/a	n/a	Process, Object, People
Halper and Krishnan (2014)	Business	8	Data management, Analytics, Infrastructure, Governance, Organisation (strategy), Skill sets, Cultural and political issues.	5	1. Nascent, 2. Pre-adoption, 3. Early-adoption, 4. Corporate Adoption, 5. Mature/Visionary	Questionnaire	n/a	n/a	Process, Object, People

Howson, (2015)	Business	5	Business drivers, People, program management, Processes, Platform-Data, BI and Analytics tools and technology.	5	1. Unaware, 2.Opportunistic, 3. Standards, 4. Enterprise, 5. Transformative	Questionnaire	n/a	n/a	Process, Object, People
Spruit and Sacu (2015)	Academic	6	1) DW Technical Solution (a. Architecture, b. Data Modelling, c. ETL, d. BI Applications), 2) DW Organisation and Processes (a. Development Process, b. Service Process)	5	1) Initial, 2)Repeatable, 3) Defined, 4) Managed, 5) Optimized	Questionnaire		Design research approach	Process, Object, People
Radcliffe (2014)	Business	8	Vision, Strategy, Value & metrics, Governance, trust & privacy, People & organization, Data sources, Data management, Analytics & visualization	6	In the Dark, Catching Up, First Pilot, Tactical Value, Strategic leverage, Optimize & Extend.	Textual-Descriptive	n/a	n/a	Process, Object, People
Knowledgent (2014)	Business	5	Business Need, Technology Platform, Operating Model, Analytics, Information Management	5	5 stages of evolution from low to high. 1) Infancy, 2) Technical Adoption, 3) Departmental Adoption,	Questionnaire	n/a	n/a	Process, object

					4) Enterprise Adoption, 5) Data as a service.				
Lavalle et al.(2010)	Academic	6	Motive, Functional Proficiency, Business challenges, Key Obstacles, Data Management, Analytics in action.	3	Aspirational, Experienced, Transformed	Questionnaire	n/a	Based on survey response	Process, object
Cosic et al.(2015)	Academic	4	Governance Capability, Culture Capability, Technology Capability, and People Capability	n/a	n/a	Textual-Descriptive	RBV	Delphi study	Process, Object, People
Wang et al.(2016a)	Academic	5	1) Sustainable SCA, 2) Agile SCA, 3) Collaborative SCA, 4) Process-based SCA, 5) Functional SCA	n/a	n/a	Textual-Descriptive	n/a	Conceptu al	Process, object
Sulaiman et al.(2015)	Academic	5	Big Data management, advanced analytics, unstructured data management, policy and governance, and visualisation	5	1) Ignorance, 2) Coping, 3) Understanding, 4) Managing, 5) Innovating	n/a	n/a	Develope d based in IDC Big Data maturity model and TWDi Big Data	Process and object

								Maturity Model	
Janssen et al.(2014)	Academic	4	Technical (Open data repository and management), Organizational(Governance and procedures for data adoption and acquisition), Networking (Community of practice, stakeholder engagement, and strategy), Juridical (assessing data sources, and illegal practice of data)	5	1) Independent, 2) Ad hoc, 3) Collaborative, 4) Integrated, 5) Unified	Textual-Descriptive	n/a	n/a	Process and object
Informs (2016)	Academic / industry	3	1) Data and infrastructure (Health, Access, Traceability, analytics architecture) 2) Analytics capability (Analytic Framework, Roles and skills, Analytics services, Analytics processes), 3) Organisation (People, leadership impact, measures, and processes)	3	1) Beginning 2) Developing 3) Advancing	Online-Questionnaire	n/a	n/a	Process, object, people

Apprendix G: Research publications

Journal Papers

- Arunachalam, D. and Kumar, N., 2018. "Benefit-based consumer segmentation and performance evaluation of clustering approaches: An evidence of data-driven decision-making." ***Expert systems with applications (Published)***

- Arunachalam, D., Kumar, N. and Kawalek, J.P., 2017. "Understanding big data analytics capabilities in supply chain management: Unravelling the issues, challenges and implications for practice." ***Transportation Research Part E: Logistics and Transportation Review (Published)*).**

- Arunachalam, D., Kumar, N., Brint, A. and Cantarelli, C.C 2018. "The role of absorptive capacity in supply chain management research: a systematic literature review." ***International Journal of Operations and Production Management (to be submitted)***.

- Kumar, N., Morales, M. G., and Arunachalam, D. 2018. "Role of big data analytics in understanding the strategies adopted by automotive companies in response to changes in environmental policy: A case study of Indian automotive sector." ***International Journal of Production Research (to be submitted in Dec 2018)***

- Kumar, N., Arunachalam, D, and Morales, M. G., 2018. "Currency Supply Chain Disruption and Business Failures: An Exploratory Study of the Demonetisation Policy in India through the lens of Big Data." ***Journal of Business Research (to be submitted in Dec 2018)***

- "The impact of Big Data Analytics practice on innovation and operational performance", (with Kumar, N., Brint, A., Cantarelli, C.C), *Working paper*.

- "Big Data Analytics adoption trend and the Digital Divide in the UK Manufacturing sector" (with Kumar, N., Brint, A., Cantarelli, C.C), *Working paper*

Conference Papers

- Arunachalam, D., Kumar, N., Brint, A. and Cantarelli, C.C 2018, **The impact of Big Data Analytics Maturity on Firm performance: Evidence from the UK manufacturing sector**, Proceedings of the 25th EurOMA Conference, Budapest, Hungary, 24 – 26th June 2018.

- Arunachalam, D., Kumar, N., Kawalek, J.P, and Shukla, N. 2017 **Exploring big data analytics capabilities for supply chain: a systematic literature review**, Proceedings of the 24th EurOMA Conference, Edinburgh, UK, 1st – 5th July 2017.

- Arunachalam, D., Kumar, N., Kawalek, J.P, 2016 **Big Data and Analytics (BDA) in the UK manufacturing supply chain: BDA capability maturity, absorptive capacity and Supply chain performance**, Proceedings of the 30th British Academy of Management Annual Conference (BAM 2016), Newcastle, United Kingdom, 6th – 8th September 2016.

- Arunachalam, D., Kumar, N., Kawalek, J.P, 2016 **Exploring the impact of Big Data Analytics capabilities: Unravelling the issues, challenges and implications for adoption and practice**, Proceedings of the White Rose Business and Management Annual Conference (WR DTC 2016), Leeds, United Kingdom, 7th – 8th July 2016