

**Gordon Institute
of Business Science**
University of Pretoria

Toward the Influence of the Organisation on Big Data Analytics

Dr. Michael John Niland

Student Number: 28163886

mjniland@gmail.com

A research project proposal submitted to the Gordon Institute of Business Science (GIBS),
University of Pretoria, in partial fulfilment of the requirements for the degree of Master of
Business Administration

6th November 2017

Abstract

Big Data Analytics (BDA) capabilities are of significant interest to organisations as they are reportedly able to enhance Firm Performance (FPer). This is of particular interest given the ever increasing demands on business to remain competitive within a dynamic market. Though BDA has been shown to be effective across diverse industries and applications, it has also been shown to be ineffective in many occasions. This study draws on theories concerning organisational capabilities, including dynamic capability theory, decision making theory and organisational culture theory to assess the influence of the organisation on the effectiveness of a BDA capability. In so doing, this study extends the above research streams by assessing the direct and moderated influences of Firm Dynamic Capability (FDC), Organisational Culture to BDA (OCDA) and Data Driven Decision Making (DDDM) on the BDA capability, and the resulting impact on FPer. The hierarchical research model was assessed with 80 online survey responses from respondents primarily based in South Africa who were directly associated with a BDA capability in their organisation. The results illustrate the significant direct and moderating impact of on BDA and FPer. The implications of these findings are discussed relative to both theoretical and practical applications in business settings, followed by considerations for further research required in the field.

Keywords

Big Data Analytics, Organisational Culture, Dynamic Capability, Data Driven Decision Making

Declaration

I declare that this research project is my own work. It is submitted in partial fulfilment of the requirements for the degree of Master of Business Administration at the Gordon Institute of Business Science, University of Pretoria. It has not been submitted before for any degree or examination in any other University. I further declare that I have obtained the necessary authorisation and consent to carry out this research.



Dr Michael John Niland

06 November 2017

Contents

Abstract.....	ii
Keywords.....	ii
Declaration.....	iii
List of Figures	viii
List of Tables.....	ix
List of Abbreviations	xi
Chapter 1: Introduction.....	1
1.1. Research problem	1
1.2. Validation of the need for this research topic	3
1.3. Business need for the study	5
Chapter 2: Literature Review	6
2.1. Introduction	6
2.2. The Big Data Environment	7
2.2.1. Overview of Research in the field of Big Data.....	7
2.2.2. Perceptions of the Big Data environment.....	8
2.2.3. Evolution of the field of Big Data and Big Data Analytics (BDA)	9
2.2.4. Dimensions of data in Big Data	10
2.2.5. Defining Big Data.....	11
2.2.6. Big Data as a value driver in business	12
2.2.7. Common challenges in the Big Data environment	15
2.3. Business Maturity.....	18
2.3.1. Maturity models in the context of this study.....	18
2.3.2. Big Data maturity models.....	19
2.4. Business Operating Models.....	20
2.4.1. Description of the business model concept.....	20
2.4.2. Business model development and innovation.....	21
2.5. Inhibitors to Implementation of Innovation in Business	21
2.6. Assimilation of Innovations and Capabilities into Organisations.....	23
2.6.1. Organisational culture toward new capabilities and innovation.....	24
2.6.2. Absorptive Capacity of Organisations	27
2.6.3. Dynamic Capabilities of Organisations.....	30
2.7. Assimilation of Innovations and Capabilities in Organisations	34

2.7.1.	Organisational Decision Making.....	35
2.7.2.	Big Data and Data Driven Decision Making.....	35
Chapter 3: Research Questions		37
3.1.	Research Overview.....	37
3.2.	Research Questions.....	38
3.2.1.	Research Question 1	38
3.2.2.	Research Question 2	38
3.2.3.	Research Question 3	39
3.2.4.	Research Question 4	39
Chapter 4: Research Methodology		42
4.1.	Introduction	42
4.2.	Research Methodology and Design	42
4.2.1.	Research Philosophy and Approach.....	42
4.2.2.	Research Type and Strategy.....	43
4.3.	Population	44
4.4.	Unit of Analysis.....	45
4.5.	Measurement by Survey Questionnaire	45
4.5.1.	Survey questions	46
4.5.2.	Measurement Instrument.....	47
4.6.	Data Collection	48
4.6.1.	Data capture by online survey	48
4.6.2.	Sampling method	48
4.6.3.	Survey distribution	50
4.7.	Survey Questionnaire Pre-testing	51
4.7.1.	Questionnaire screening	51
4.7.2.	Sample respondent pre-test	52
4.8.	Survey Data Collation	52
4.8.1.	Background on managing missing survey data	53
4.8.2.	Management of missing survey data in this study	54
4.8.3.	Data coding for imputation	55
4.8.4.	Imputation of missing survey data by averaging	55
4.8.5.	Final Sample Size	55
4.9.	Analysis Approach	56

4.9.1.	Descriptive statistics.....	56
4.9.2.	Internal reliability of data and constructs (Cronbach's alpha)	57
4.9.3.	Internal validity of data and constructs	58
4.9.4.	Factor Analysis (FA) of data.....	59
4.10.	Assessment of the relationship between constructs.....	62
4.10.1.	Correlation between constructs	62
4.10.2.	Assessment of the relative influence of constructs on firm performance through hierarchical multiple regression (HMR)	63
Chapter 5: Results and Analysis		66
5.1.	Introduction	66
5.2.	Descriptive Characterisation of the Sample of Data	66
5.2.1.	Raw sample size	66
5.2.2.	Total data sample.....	67
5.2.3.	Normality of data per survey question	69
5.2.4.	Descriptive characteristic of respondents and their organisations	69
5.3.	Assessment of Survey Sections as Constructs.....	74
5.3.1.	Internal reliability of data sets (Cronbach's alpha)	75
5.4.	Factor Analysis of constructs	75
5.4.1.	Exploratory Factor Analysis.....	75
5.4.2.	Confirmatory Factor Analysis	77
5.5.	Construct Descriptive Statistics.....	82
5.6.	Assessment of Research Hypotheses.....	83
5.6.1.	Correlation of research constructs.....	83
5.6.2.	Hierarchical multiple regression of research constructs	87
Chapter 6: Discussion of Findings		90
6.1.	Introduction	90
6.2.	Assessment of the relationship of constructs to FPer	90
5.7.1.	Discussion of Research Question 1	90
5.7.2.	Discussion of Research Question 2	91
6.3.	Assessment of the relationship of constructs to BDAC	94
5.8.1.	Discussion of Research Question 3	94
6.4.	Assessment of the relative weighting and impact of constructs on the BDA – FPer relationship	98
6.4.1.	Discussion of Research Question 4	98

6.5.	Assessment of BDA in the business context	102
6.5.1.	Apprehension to BDA by people in the business environment as a result of a lack of understanding.....	102
6.5.2.	Engagement by BDA with business to improve output	104
Chapter 7: Conclusion		106
7.1.	Introduction	106
7.2.	Contributions made toward theory by the research study	106
7.3.	Practical implications that can be relayed to business	108
7.4.	Recommendations for future research.....	109
7.5.	Limitations of research.....	111
References		115
Appendices		137
9.1.	Appendix 1: Research Questionnaire.....	137
9.1.1.	Introduction	137
9.1.2.	Questionnaire.....	138
9.1.3.	Additional survey distribution data.....	143
9.1.4.	Challenges of data collection by online survey	143
9.2.	Appendix 2: Ethical Clearance	145
9.3.	Appendix 3: Data Preparation.....	146
9.3.1.	Data Collected	146
9.3.2.	Data Coding	146
9.3.3.	Data Imputation	148
9.4.	Appendix 4: Data Analysis.....	154
9.4.1.	Descriptive analyses of data.....	154
9.4.2.	Assessment of internal reliability (Cronbach's alpha).....	156
9.4.3.	Assessment of construct internal validity by Pearson's correlation	158
9.4.4.	Assessment linearity between variables by scatter plot for Factor Analysis....	160
9.4.5.	Exploratory factor analysis (EFA)	162
9.4.6.	Confirmatory factor analysis	164
9.4.7.	Correlation of Factors	169
9.4.8.	Scatter plots of each factor (Y) and FPer (X)	172

List of Figures

Figure 1: Base research model applied in the study.....	41
Figure 2: Number of Missing Answers per Question and Section of the Questionnaire	68
Figure 3: Revised research model applied in the study	81
Figure 4: Representative scatter plots illustrating linearity of variables across constructs (FDC, OCBDA and DDDM)	160
Figure 5: Scree plot of sections 2 (OCBDA)	163
Figure 6: CFA model of OCBDA	164
Figure 7: CFA of DDDM	165
Figure 8: CFA of FDC and FAC.....	165
Figure 9: CFA of BDAMC.....	166
Figure 10: CFA of BDAIC	166
Figure 11: CFA of BDAPE	167
Figure 12: CFA of FPer	167
Figure 13: CFA of the third order construct BDAC	168
Figure 14: BDAC loading of level 2 constructs	168
Figure 15: Scatter plot of each factor with FPer to illustrate relative distribution.....	172

List of Tables

Table 1: Three tiers of value creation as a result of Big Data	14
Table 2: Summary Descriptive statistics of survey constructs	58
Table 3: Summary of data collected and data imputed for application in analyses.....	68
Table 4: Summary of missing or imputed data per section of the research survey	69
Table 5: Country in which respondents work	70
Table 6: Industries represented by survey.....	71
Table 7: Number of employees within respondents organisations.....	72
Table 8: Years for which respondents organisations have actively pursued BDA	72
Table 9: Current job level of respondents	73
Table 10: Main association with the BDA capability.....	73
Table 11: Years of work experience in the field of BDA.....	74
Table 12: Summarised reliability statistics of internal sample data (Cronbach's alpha)	75
Table 13: Summary results from Exploratory Factor Analysis to identify factors from survey constructs.....	76
Table 14: Summary results from Confirmatory Factor Analysis per construct.....	77
Table 15: Descriptive statistics of factors extracted from EFA and CFA	82
Table 16: Assessment of factor normality through Shapiro-Wilk testing.....	83
Table 17: Correlation coefficients of constructs with Firm Performance.....	84
Table 18: Correlation coefficients of constructs with BDA capability.....	86
Table 19: Summary of heirarchical regression analysis of factors (FDC, OCBDA and DDDM) impacting on the BDAC - FPer relationship.....	88
Table 20: Summary correlations and contribution to variance by 1 st order and 2 nd order constructs.....	92
Table 21: Survey Questionnaire and Answer Options	138
Table 22: Forums used for Survey Distribution.....	143
Table 23: Sample extract of data collected using SurveyMonkey (only a portion of respondents and only data up to Section 1 are illustrated).....	146
Table 24: Coding of English language variables to numeric values	147
Table 25: Illustrative results of coding processes	148
Table 26: Summary results of data collected and data imputed	149
Table 27: Selection of grouping for imputation of a missing answer in the highlighted cell....	150
Table 28: Complete Coded and Imputed Data Set.....	151

Table 29: Assessment of descriptive statistics for each question.....	154
Table 30: Example of Cronbach alpha item-total statistics to illustrate which items to delete to increase total reliability score	157
Table 31: Representative Cronbach's alpha assessment of Section 3 (DDDM) of the survey ..	157
Table 32: Confirmation of internal validity per construct through two Pearson's correlation	158
Table 33: KMO and Bartlett's test of Sections 2 (OCBDA)	162
Table 34: Extraction of number of constructs and total variance explained by Eigenvalues for Section 2 (OCBDA).....	162
Table 35: Rotated Component matrix illustrating factor loading per item from EFA for Section 2 (OCBDA)	164
Table 36: Spearman's correlation matrix between all construct factors assessed.....	169
Table 37: Regression of 2 nd order BDAC constructs to BDAC (3 rd order)	171

List of Abbreviations

- AWS – Amazon Web Services
- B – Byte (1B)
- BDA – Big Data Analytics
- BDAC – Big Data Analytics Capability
- BDAIF – BDA Infrastructure Flexibility
- BDAMC – BDA Management Capability
- BDAPE – BDA Personnel Expertise
- BI – Business Intelligence
- CDR – Call Data Record
- CFA – Confirmatory Factor Analysis
- HMR – Hierarchical multiple regression
- IFI – Incremental fit index
- IoT – Internet of Things
- IP – Intellectual Property
- IS – Information Systems
- IT – Information Technology
- KMO – Kaiser-Mayer-Olkin index
- OCBDA – Organisational Culture to Big Data Analytics
- CFI – Comparative fit index
- CV – Coefficient of Variation
- DCT – Dynamic Capability Theory
- DDDM – Data Driven Decision Making
- EFA – Exploratory Factor Analysis
- FAC – Firm Adsorptive Capability
- FDC – Firm Dynamic Capability
- FPer – Firm Performance
- GB – Giga Byte (10^9 B)
- HLR – Hierarchical linear regression
- POPI – Protection of Personal Information Act of South Africa
- RMSEA – Root mean square error of approximation
- SA – Republic of South Africa
- SEM – Structured equation Modelling
- TB – Tera Byte (10^{12} B)
- ZB – Zetta Byte (10^{21} B)
- α – Cronbach's alpha
- λ – Factor loading

Chapter 1: Introduction

“It ain’t what you know that gets you into trouble. It’s what you know for sure that just ain’t so.” – Mark Twain

1.1. Research problem

As a result of the vast promise that Big Data Analytics (BDA) holds for businesses from diverse industries to become more competitive, it has received ever increasing attention from across mainstream media, the business sector, and academia since 2011 (Chen, Chiang, & Storey, 2012; Chen, Mao, & Liu, 2014; Comuzzi & Patel, 2016; Frizzo-Barker, Chow-White, Mozafari, & Ha, 2016; Hashem *et al.*, 2015; Khan, Liu, Shakil, & Alam, 2017; Larson & Chang, 2016; Lavalley, Lesser, Shockley, Hopkins, & Kruschwitz, 2011; Lee, 2017; McAfee & Brynjolfsson, 2012; Özköse, Ari, & Gencer, 2015; Wamba, Akter, Edwards, Chopin, & Gnanzou, 2015). ‘Big Data’ is a collective term coined to describe group of specialised technologies and capabilities which together enable the collection, storage, processing and targeted application of immense amounts of data (Chen *et al.*, 2012, 2014; Hashem *et al.*, 2015; Özköse *et al.*, 2015). BDA is the analytics output from the application of statistical methods and computer aided analysis to these very large data by practitioners in the field of ‘Data Science’ (Chen & Zhang, 2014; Davenport & Patil, 2011; Larson & Chang, 2016; McAfee & Brynjolfsson, 2012; Provost & Fawcett, 2013). In order to enable the practitioners field of data science to achieve the potential of BDA, the Big Data environment and its supporting systems, capabilities, technologies and functions have undergone rapid development since 2011 (Chen *et al.*, 2012, 2014; Gil & Song, 2015; Hashem *et al.*, 2015; Lee, 2017; Özköse *et al.*, 2015; Shin, 2016). Given both the newness of the Big Data environment, and the significant requirement to develop the technical capabilities required to enable it, this has been the focus of research activities, with up to 87% of research in the field being technical in nature (Alharthi, Krotov, & Bowman, 2017; Chen & Zhang, 2014; Frizzo-Barker *et al.*, 2016; Pospiech & Felden, 2012; Wamba *et al.*, 2015). The result of this is that despite BDA being a capability which must be assimilated into businesses for it to deliver value, this remains an under researched area of literature (Akter, Wamba, Gunasekaran, Dubey, & Childe, 2016; Alharthi *et al.*, 2017; Frizzo-Barker *et al.*, 2016; Mazzei & Noble, 2017; Wamba *et al.*, 2017, 2015). This research is accordingly directed at

further understanding the impact of the organisation as an environment into which BDA must assimilate, on the effectiveness of BDA as a functional capability. The intent of this research was to further illuminate the characteristics of this interaction such that practical and actionable recommendations could be suggested for managers and senior leadership of organisations which intend to develop BDA capabilities as a vehicle to deliver sustained value and competitiveness to the business.

The concept of value being created through data analytics is by no means a new phenomenon. Conclusive verification of what this approach can deliver to business has been reported in numerous studies (Brynjolfsson, Hitt, & Kim, 2011; Chen *et al.*, 2012; Kiron, Prentice, & Ferguson, 2014; Kiron & Shockley, 2011; McAfee & Brynjolfsson, 2012; Williams, 2003). Though validation for the potential of BDA, and the need to apply it effectively appears evident, it appears that this is a challenging task for organisations with many technological, legal and business environment complexities (Chen & Zhang, 2014; Chen *et al.*, 2012; Frizzo-Barker *et al.*, 2016). Though often reported as being a function that can add significant value to business through processes such as improving efficiencies, developing new capabilities and products and increasing competitiveness, limited literature that deals with the deployment of BDA capabilities in an organisational context (Alharthi *et al.*, 2017; Frizzo-Barker *et al.*, 2016; Wamba *et al.*, 2015). There is thus a need to improve the understanding of the parameters which impact on the deployment of BDA capabilities within the environment of an organisation, a though many organisations are attempting to establish BDA, and some report pockets of capability, there are a limited number of effective deployments in market (Akter *et al.*, 2016; Alharthi *et al.*, 2017; Comuzzi & Patel, 2016; Frizzo-Barker *et al.*, 2016; Mazzei & Noble, 2017; Wamba *et al.*, 2017, 2015).

Due to the internet and access via personal computers and mobile devices, information is far more freely available today than ever before. To give representation to scale of current data availability, the entire printed content of the Library of Congress, the biggest single collection in the world, can be stored on 1 Exabyte (10^{18} bytes) 100,000 times over (Hirsch, 2017). In the year 2016 alone, information to the tune of 1,100 Exabytes was shared via the internet (Hirsch, 2017). The key issue which the body of Big Data literature has been addressed since 2011 has been how to collect, process, curate and distil these data to actionable insight. Key developments which have enabled the BDA capability are that firstly, the cost of storage has declined from 100's of thousands of US\$ per GB in the 1980's to US\$ 0.02 per GB and below in 2016 (Coughlin, 2016; Komorowski, 2009; LaChapelle, 2016). Secondly, in accordance with

Moore's law, computing power (ability to process data) has doubled every 18 - 24 months since the 1980's, which resulted in an exponential increase in processing capability to date (Hirsch, 2017; Schaller, 1997; Simonite, 2016). And thirdly that software and networking have developed to the point that it has enabled virtual deployment of infrastructure to manage these processes in parallel, allowing for all the data to be 'read' at the same time (Hirsch, 2017). These technical developments describe one half of the journey from data to value. The second half involves the directed assessment of the data (Data Science), extraction of insight that can be actioned, and then effectively applying that in an organisation such that it can benefit from it (Davenport & Patil, 2011; Debortoli, Müller, & Vom Brocke, 2014; Fang, Jiang, & Song, 2016; Fong, Fang, & Luo, 2015; Zaki & Meira, 2013; Zhou, Pan, Wang, & Vasilakos, 2017).

Given the reported capabilities of BDA, it is positioned to make significant impacts in business if it can be properly executed so that it delivers competitive advantage (Chen *et al.*, 2014; Lavallo *et al.*, 2011; Mazzei & Noble, 2017). The execution of BDA within business appears to be a current shortfall, thus it is the focus of this study.

1.2. Validation of the need for this research topic

Top performing organisations are twice as likely to apply analytics in their business processes (Lavallo *et al.*, 2011). Brynjolfsson *et al.* (2011) performed a detailed survey and found that companies that were data driven tended to have output and productivity that was 5% - 6% higher than what would be expected given their other investments. Similarly, Wamba *et al.* (2015) reported that a 15% - 20% ROI, whilst Nannetti (2012) found that of their sample, performance had improved by 26% through big data. Furthermore, McAfee and Brynjolfsson (2012) reported that the more companies characterise themselves as data-driven, the greater the propensity for them to both outperform their peers in market. With the range and depth of application has been shown for BDA applications, it now maintains impact in virtually all business types, industries and disciplines (Brown, Chui, & Manyika, 2011; Chen *et al.*, 2012, 2014; Colas, Finck, Buvat, Nambiar, & Raj Singh, 2014; Fan, Lau, & Zhao, 2015; Hashem *et al.*, 2015; Lavallo *et al.*, 2011; Siddiqua *et al.*, 2016; Yaqoob *et al.*, 2016).

All of these studies describe a high will and reported appreciation for the need to apply BDA, achieving the potential of these capabilities has however been limited to date (Bradley, Barbier, & Handler, 2013; Buhl, 2013; Wang, Xu, Fujita, & Liu, 2016). The successes of BDA applications are typically confined to in single projects or pockets within businesses and are

thus not descriptive of capabilities which deliver sustained competitive advantage and provide input to strategic directives (Elgendy & Elragal, 2016; Espinosa & Armour, 2016; Lee, 2017; Mazzei & Noble, 2017; McAbee, Landis, & Burke, 2016; Power, 2015; Wang *et al.*, 2016). This is evidenced through only half a percent of the 'big data' that organisations have available is, on average, being analysed for insight (Bradley *et al.*, 2013), and up to half of all BDA projects in large organisations are never completed (Comuzzi & Patel, 2016).

Though BDA is an actively discussed topic in literature due to the significant potential which promises (Alharthi *et al.*, 2017; Bendle & Wang, 2016; Chen *et al.*, 2012; Fan *et al.*, 2015; Gil & Song, 2015; Habib, Chang, Batool, & Ying, 2016; Hashem *et al.*, 2015; Özköse *et al.*, 2015; Shah, Irani, Sharif, & Job, 2017; Siddiqua *et al.*, 2016; Wamba *et al.*, 2015; Y. Wang & Hajli, 2017; Wang *et al.*, 2016) it is most often poorly comprehended by business people. By nature, BDA projects are broad reaching as data are not only generated at all levels of an organisation, but the outputs of BDA programs should be implemented across multiple divisions and decision making levels in an organisation (Chen *et al.*, 2014; Comuzzi & Patel, 2016; Wang *et al.*, 2016). This thus requires broad cultural appreciation, understanding and acceptance of BDA as a capability by the organisation (Alharthi *et al.*, 2017) in order for it to achieve its potential.

Lavalle *et al.* (2011) asserted that among the leading causes of poor BDA implementation are organisational dimensions of lack of understanding of how to use analytics to improve the business, lack of focus by management with competing priorities, insufficient business processes for both data and implementation and organisational culture unreceptive to change. Though advanced technical capabilities will remain scarce due to the rate at which the technology advances, in most cases the base capabilities for BDA are relatively easily attainable, particularly with outsourcing of technical requirements, cloud based storage and computing being readily available and online training materials being readily available. The benefits of big data processing are contested by the ability to use it effectively as a result of limited clarity on how to achieve this by general business people. This is due to it being a rapidly developing space at the interface of several complex fields which is still in its infancy (Chen *et al.*, 2012; Fan *et al.*, 2015; Gil & Song, 2015; Hashem *et al.*, 2015; Özköse *et al.*, 2015; Siddiqua *et al.*, 2016; Wamba *et al.*, 2015). The need for effective processes to identify, extract and implement sustained value from big data platforms is a highly relevant and topical field with potentially significant business applications (Alharthi *et al.*, 2017; Chan, 2014; Chen *et al.*, 2012; Erevelles, Fukawa, & Swayne, 2016; Fan *et al.*, 2015; Hashem *et al.*, 2015; Kshetri, 2014; Lavalle *et al.*, 2011; Özköse *et al.*, 2015).

1.3. Business need for the study

This study is necessary as it adds to the, to date, under researched area of assessing factors which influence a business's capability to use, implement and capitalise on BDA as a tool. This is undertaken in the light of the clear potential of such capabilities and the rapid advancement of the field in business. The business need is thus to leverage value from BDA through understanding how to use the capability and what the critical success attributes are.

More specifically, this study is directed at assessing three dimensions of organisational capability and their impact on the BDA, in order to ascertain the degree of influence which they have on BDA, and as a result, firm performance. This is necessary as a majority of research focuses on development of the systems, infrastructure and the current strategies driving systems and processes (physical, managerial and technological) that enable value (market share, efficiency, profit and sustained competitiveness) to be realised from big data. Given the highly mathematical and technical nature of the big data environment, the available literature base appears to be out of the appreciable reach of a majority of business people. Unfortunately, these are the very people who appear to be hindering the current application of big data platforms as they are not actively driving appropriate business models innovations to effectively incorporate these platforms.

The objective of this study is to identify how the factors influence the outcome of BDA initiatives, such that they can be more effectively packaged and presented for consumption by business people, and hence allow for better implementation. This study thus intends to move toward a better understanding of what is necessary to facilitate the application of a BDA capability, and to provide tangible and achievable recommendations to business to support delivery of sustained competitive advantage.

Chapter 2: Literature Review

“Knowledge is Power” – Francis Bacon

2.1. Introduction

The contents of this literature review are directed at illustrating the development of the Big Data environment, and outlining the current outlook on its impact on business. The current state of research is considered in three broad areas, including firstly aspects of the Big Data environment (the actual processing of enormous data sets). Secondly, the impact of Big Data *on* its environment (impacts observed from application of insight from analysis of these data). Finally, importantly in the context of this study, the impact of the environment *on* Big Data (whether external factors impact on its efficacy) is considered. In so doing, this review intends to describe the necessity and basis for this study, as it is intended to enable improved realisation of value from Big Data Analytics (BDA) capability.

According to the above subdivisions, a review of current literature is presented, describing the Big Data environment with a focus on descriptive characteristics of each. This is done in order to position the current state of knowledge surrounding the environment and the gaps and opportunities which exist; describing the research of which this thesis is comprised. Given that the application of BDA is a capability which remains in relative infancy in most cases, it is considered to be one still in development. This is contrasted with business as an environment, and considers the characteristics of that environment and how the environment interacts with BDA in order to describe required characteristics for effective implementation of BDA in business. Particular focus was paid to factors influencing the relationship, with the intent of describing how to enable business to realise the sustained value and competitive advantage that is the potential of this environment.

The theory bases of organisational culture toward innovation, decision making theory dynamic capability theory and absorptive capability theory are applied as constructs in the assessment of whether Big Data capabilities and analytics derived from them perpetuate or support improved firm performance.

2.2. The Big Data Environment

Digital technologies have influenced, and in many cases governed, the way in which organisations are built, how functions are maintained and how their future actions are directed (Elgendy & Elragal, 2016). Recent years have seen significant interest in the field of Big Data as a technology, and as a capability through BDA. This is both as a means of achieving and attaining competitive advantage, and as a source of information which may direct future business across virtually any industry or discipline (Brown *et al.*, 2011; Chen *et al.*, 2012, 2014; Colas *et al.*, 2014; Fan *et al.*, 2015; Hashem *et al.*, 2015; Lavalley *et al.*, 2011; Siddiqua *et al.*, 2016; Yaqoob *et al.*, 2016). The extent of recent papers and reviews concerning Big Data and its applications illustrate the significant interest towards resolving challenges and misconceptions currently faced within the Big Data arena (Akter *et al.*, 2016; Alharthi *et al.*, 2017; Boyd & Crawford, 2012; Chen & Zhang, 2014; Chen *et al.*, 2012; Fan *et al.*, 2015; Gil & Song, 2015; Hashem *et al.*, 2015; Jagadish, 2015; Kambatla, Kollias, Kumar, & Grama, 2014; Lee, 2017; Liu, Li, Li, & Wu, 2016; Malthouse, Haenlein, Skiera, Wege, & Zhang, 2013; Ram, Zhang, & Koronios, 2016; Shin, 2016; Wamba *et al.*, 2017, 2015).

The following section is intended to describe the development of the Big Data environment and characterise both the current capabilities and the gaps present, in order to position it as a capability within business.

2.2.1. Overview of Research in the field of Big Data

Pospiech and Felden (2012) clustered Big Data literature into technical and functional data with provisioning and utilisation buckets respectively. They found that 87% of literature is technical in nature. This describes that a majority of literature available focusses on the technical attributes of managing, collecting, processing and analysing data itself. This is a necessary feature as the big data environment is based on very recent, and constantly developing technical capabilities, with many avenues which still require significant attention. These include, among others, what data to use, what weighting to apply to data depending on application and how to integrate and manage heterogeneity between data sources and skewness of data sets and importantly how to manage sparsely populated data fields, and then how to curate these data such that they can be accessed and factorised for application in modelling, to name but a few (Erevelles *et al.*, 2016; Fan *et al.*, 2015; Hashem *et al.*, 2015). Subsequent to that, when actually processing the data, considerations such as what joins to make between data sets, what analytical systems to use, what modelling techniques to apply,

what algorithms and structures to use in the processing of data given data types, volumes and applications remains of significant area of interest within literature, with ever advancing developments (Bhattacharya, Islam, & Abawajy, 2016; Chen & Zhang, 2014; Gandomi & Haider, 2015; Gil & Song, 2015; Kambatla *et al.*, 2014; Miloslavskaya & Tolstoy, 2016; Nair, Misra, Hornbuckle, Mishra, & Acharya, 2014).

The relatively small proportion of functional papers described by Pospiech and Felden (2012) may provide a basis for why the field appears to be poorly understood by a majority of business people, resulting in a lack of clarity in boardrooms regarding what to make of the data available or how to apply it within the business (Elgendy & Elragal, 2016; Espinosa & Armour, 2016; Lismont, Vanthienen, Baesens, & Lemahieu, 2017; Mazzei & Noble, 2017).

Given the lack of conformity and clarity on how to appropriately apply BDA in business, and how to leverage these capabilities, further research into business processes, models and systems are required to better identify influencing factors and how to action insights (Akter *et al.*, 2016; Bommel, Edelman, & Ungerman, 2014; Brown *et al.*, 2011; Colas *et al.*, 2014; Lavallo *et al.*, 2011; Mazzei & Noble, 2017; Nannetti, 2012; Wamba *et al.*, 2017). This arena describes a business rather than a technical focus, and is core area of concern in this study.

2.2.2. Perceptions of the Big Data environment

Application of Big Data technologies, as internal business capabilities, and as supplied services, is forecast to grow significantly in the near future. An average compounded annual growth rate of 23.1% is expected between 2014 and 2019, with a corresponding increase in annual spend to US\$ 48.6 billion in 2019, from a point where it was barely considered in the late 2000's (Chen *et al.*, 2014; Lee, 2017). This is a feature that stems from a broad perception in market that BDA has the potential to significantly improve the competitive advantage of a business. Though there appears to be appreciation for the potential, there is a distinct perception that BDA is extremely complex, and out of the conceptual reach of the average person in business.

The following sections intend to discuss technical characteristics of the Big Data environment with the intent of describing these aspects in a palatable manner. This is done with the intent of demystify the capability, and actively bringing it to a point where it can be understood by people in business who have the influence to make business decisions and influence structures and processes. If this is done, it is anticipated that BDA may be viewed as a tool from a business perspective, rather than a complex analytical 'black box' only appreciated by the data

scientists. It seems that Data Scientists, the technicians of BDA, may be considered in a light not dissimilar to how Rocket Scientists were in years gone by. That is to say, Data Scientists are immediately assumed to be remarkably intelligent, in that they decipher the strange and wonderful world of Big Data and bend it to their will. It is the strong position of the author that this is a disservice both to BDA as a capability and to business in general, as it distances the average person in business from trying to understand BDA, and hence how to use or implement it as a tool in business. It is therefore essential to understand this environment and its capabilities so that it can be effectively explained to the average person in business in order for it to be embraced as a functional tool, rather than a distant potential.

2.2.3. Evolution of the field of Big Data and Big Data Analytics (BDA)

The field of Big Data is relatively new, and is one which has been enabled by the development of technologies including networking (allowing collection of data), hardware (vastly improved computing capability), storage (dramatically reduced cost in hard drive) as well as software (programs that allow for data to be analysed) (Chen & Zhang, 2014; Chen *et al.*, 2012; Lee, 2017; Özköse *et al.*, 2015; Pospiech & Felden, 2012; Siddiqi *et al.*, 2016). These capabilities have largely been developed and driven by private companies such as IBM, Microsoft, Google, Amazon and Facebook, to name but a few, who needed to develop systems and capabilities to facilitate their business needs (Chen *et al.*, 2014).

According to Lee (2017), the first stage in the evolution of Big Data began as early as 1994, and was in order to manage web content and mine the usage data to determine user behaviour. Chen *et al.* (2014) reported that as early as 2001, the opportunities and challenges which arose as a result of the 3V's (Velocity, Variety and Volume, described further below) were being considered.

The next phase of evolution arose as a result of the development of Web 2.0 in 2005, which allowed users to begin to interact with web-sites and contribute content to those sites, giving rise to social media and the rise in business interest in the arena (Lee, 2017; Provost & Fawcett, 2013). By 2007 this drove the need to develop a new generation of computing tools to manage, process, analyse and visualise these already challengingly large data sets (Chen *et al.*, 2014). Though Big Data began gaining more general notoriety in 2008 when Nature published a special issue on Big Data, it has only seen a significant increase in application since 2011 (Lee, 2017). In 2012 it was announced at the Davos Forum in Switzerland that Big Data had become a new form of 'economic asset' (Chen *et al.*, 2012; Lee, 2017; Pospiech & Felden,

2012). Lee (2017) reported that the third stage of Big Data's evolution began in 2015, led by the development of the Internet of Things (IoT) and the tremendous increases in both data and opportunities that are associated with interconnectivity of devices.

Understandably, Lee's (2017) classifications are determined by technical delineations that follow the rapid technological and analytical developments that are primarily discussed in literature. Functional and practical applications of capabilities in business are however comparatively underweight in literature. Rather, as is described by Akter *et al.* (2016), Big Data tends to be generally classified as having the 'potential to transform management theory and practice', to be the 'next big thing in innovation' or that it will drive the 'next management revolution'. Unfortunately, as is common in literature, description of how this should or could actually be implemented or classified is limited.

Towards this need, recent literature has begun to table maturity models and business models for the Big Data environment, both of which are discussed later (Accenture Digital, 2015; Comuzzi & Patel, 2016; Domino Data Lab, n.d.; Halper & Krishnan, 2014; Hortonworks, 2016). Typically however, applications of systems seem to be bespoke to tasks within organisations, and do not seem to have reached consistent and sustained application across business as yet. That said, the rapid growth and acceptance of platforms such as Amazon Web Services (AWS) and Microsoft Azure, which facilitate Big Data processes through cloud infrastructure, is evidence of increased acceptance and implementation of Big Data capability in business today.

2.2.4. Dimensions of data in Big Data

The past 30 years have seen an unprecedented increase in the production of, and access to, data by business in huge volumes, varieties and velocities (Chan, 2014; Chen *et al.*, 2012; Erevelles *et al.*, 2016; Fan *et al.*, 2015; Kshetri, 2014; Özköse *et al.*, 2015; Wamba *et al.*, 2015). *Velocity* describes the speed at which data is produced and the timeliness required to collection and analyse that data. *Variety* speaks to the many types of data which are collected, ranging from structured (transaction details, times, dates, names, locations, etc.) to unstructured data (audio, video and free text content, etc.). *Volume* describes the sheer amount of data generated and collected, which continues to become increasingly big (Chen *et al.*, 2014; Elgendy & Elragal, 2016; Emani, Cullot, & Nicolle, 2015; Lee, 2017; Pospiech & Felden, 2012). In addition to these core features, further V's have been proposed. These include firstly '*Value*' to emphasise the importance of using data to increase revenue (Elgendy & Elragal, 2016; Lee, 2017). Finally, *Veracity* describes the uncertainty and unreliability which

may arise due to incompleteness, inaccuracy, latency, inconsistency and subjectivity of data sets (Elgendy & Elragal, 2016; Lee, 2017).

At a global scale, the rate and extent of data being produced is forecast to continue to increase at a nearly incomprehensible rate. Chen *et al.* (2014) reported that in 2011 around 1.8ZB (10^{21} B) of data were created and copied. Though it took nine years for volume to double to this point, they report further that this volume is expected to double at least biannually into the near future. This is in part driven by social media and the more recent phenomenon of the IoT, with 4.9 billion connected objects in 2015 (up 30% from 2014) expected to reach 25 billion by 2020 with 5.5. million new devices connected daily (Lee, 2017). Emani *et al.* (2015) report that it is forecasted that by 2025, the data produced on the Internet alone will exceed the brain capacity of every individual in the entire world. The combined result of more devices being more connected through the IoT, as well as the extent of ever expanding social media content is persistently expanding the availability of data to business.

Though typical businesses would have no direct application for the total scale of data, the scale and rate of increase of data is relevant at the scale of typical organisations, as their own business operation, customers, equipment and sales are producing data at the same rate of increase as mentioned on a global scale. It is this scale of data that BDA applications need to be able to make use of, and is these data that organisations need to find a way of leveraging in order for them to achieve sustained competitive advantage.

2.2.5. Defining Big Data

As may be expected of a young and rapidly advancing field, particularly one being developed by separate players in industry with sometimes limited congruence in terminology, there is limited consensus in the definition of the term 'Big Data'. This is also an artefact of 'Big Data' being a term for an ecosystem of multiple parts, each of which are developing rapidly. Özköse *et al.*, (2015) summarised definitions of Big Data in literature, and found that the most commonly ones appear to focus on the amount of data, wherein 'Big Data' tends to represent the physical or technological requirements for storage, processing, management and analytics are just beyond prevalent technology's capability. Wang, Xu, Fujita, and Liu (2016) provided a more structured approach to defining Big Data. They did this though separating out four perspectives based on limits of application Big Data;

- **The product-oriented perspective** maintains the historical perspective and focuses on the attributes of data itself as described by size, speed and structure. Herein, Big Data may be

defined as: “*Large, diverse, complex, longitudinal, and/or distributed data sets*” which are ‘big’ because “*there is too much of it, because it is moving too fast, or because it is not structured in a usable way*” (Wang et al., 2016). Though an ever increasing value, Lee (2017) reports that the minimum threshold required to classify data as ‘Big Data’ is currently 1TB, which is approximately 1,500 CDs or 220 DVDs worth of data, which is sufficient storage for 16 million Facebook photographs.

- **The process-oriented perspective** focusses on the technological tools (programs, coding languages, computational requirements, statistical and technological techniques) required to store, manage, aggregate and analyse data. Herein, Big Data may be defined as when: “*Normal application of current technology does not enable users to obtain timely, cost-effective and quality answers to data-driven questions*” or similarly when forced to “*look beyond the tried-and-true methods that are prevalent at that time*” (Wang et al., 2016).
- **The cognition-based perspective** characterises ‘Big Data’ as that which exceeds the cognitive limits of “*commonly used hardware environments and software tools to capture, manage, and process it within a tolerable elapsed time for its user population*” or which is “*beyond the ability of typical database software tools to capture, store, manage and analyse*” (Wang et al., 2016).
- **The social movement perspective** emphasises the gap between perception and reality present in underlying data. Herein, Big Data may be defined as: “*The potential to quantify and change various aspects of contemporary life, to revolutionize the art of management, or to be part of a major transformation that requires national effort.*” (Wang et al., 2016).

Though each of the above can be applied within defined circumstances, it is useful to maintain an overarching definition, to characterise the environment more broadly, and then to deal with components of the environment where specifics are required. It is for this reason that the following was selected as a base description for this study; “*to harness [extensive] information in novel ways to produce useful insights of goods and services of significant value*” (Comuzzi & Patel, 2016).

2.2.6. Big Data as a value driver in business

Literature is filled with anecdotal assertions about the potential of Big Data platforms to produce immensely valuable and impactful findings and insights. Effectively using information available to the business requires the linking of generation of insight, knowledge management and resulting operational execution to deliver firm performance. This has been illustrated directly, through increasing operational efficiencies, as well as indirectly through improved

decision making, appropriate product development or placement choices and through knowledge or information having value in its own right (Chen *et al.*, 2012; Lee, 2017; Pospiech & Felden, 2012). The application of BDA has been assessed in several market analyses (Brown *et al.*, 2011; Colas *et al.*, 2014; Kiron *et al.*, 2014; Lavalley *et al.*, 2011; Nannetti, 2012). All of these studies describe a high will and reported appreciation for the potential for BDA to improve business performance, however execution is often not effective (Bradley *et al.*, 2013; Wang *et al.*, 2016). Though significant benefits are described, these are often successes in single projects or pockets within businesses and are thus not descriptive of mature Big Data deployments in business which enables sustained competitive advantage and provide input to strategic directives (Elgendy & Elragal, 2016; Espinosa & Armour, 2016; Lee, 2017; Mazzei & Noble, 2017; McAbee *et al.*, 2016; Power, 2015; Wang *et al.*, 2016).

Not all businesses are equally capable of extracting value from their data, let alone from Big Data. There are however numerous anecdotal references to value being added significant value to firms, ranging from directly impacting on competitive capability (Espinosa & Armour, 2016; Kubina, Varmus, & Kubinova, 2015; Mikalef & Pateli, 2017; Özköse *et al.*, 2015), facilitating firms to tackle current business challenges (Kubina *et al.*, 2015; Perrons & Jensen, 2015), developing new organisational capabilities (Addo-Tenkorang & Helo, 2016; Akter *et al.*, 2016; Braganza, Brooks, Nepelski, Ali, & Moro, 2017; Crowston & Qin, 2011), enabling innovation (Brown *et al.*, 2011; Gobble, 2013; Mazzei & Noble, 2017; Vargo & Lusch, 2004), service improvement (Bommel *et al.*, 2014; Comuzzi & Patel, 2016; Huang, Zhong, & Tsui, 2015; Ostrom, Parasuraman, Bowen, Patrício, & Voss, 2015), personalising marketing and targeting consumers (Bommel *et al.*, 2014; Fan *et al.*, 2015; Laroche, Bergeron, & Barbaro-Forleo, 2001; Lee, 2017; Loebbecke & Picot, 2015; Nair *et al.*, 2014; Summers, Smith, & Reczek, 2016; Xu, Frankwick, & Ramirez, 2016; Zuiderveen Borgesius, 2016), profiling developments and product features (Du, Hu, & Damangir, 2015; Sil & Das, 2016), supply chain management (Addo-Tenkorang & Helo, 2016; Waller & Fawcett, 2013; Zhong, Newman, Huang, & Lan, 2016), improved decision making and management capabilities (Comuzzi & Patel, 2016; Horita, de Albuquerque, Marchezini, & Mendiondo, 2017; Lau, Yang-turner, & Karacapilidis, 2014; McAfee & Brynjolfsson, 2012; Monino, 2016; Sil & Das, 2016; Wang *et al.*, 2016) among others.

Given the breadth of application and the depth to which BDA can have impact on businesses across virtually all industries, the potential for value creating is impressively large. Mazzei and Noble (2017) provide a three tiered framework to conceptualise the value created by Big Data in business (Table 1), wherein as one progresses through the tiers, data becomes more a driver

of strategy than a result of strategy. In so doing, they embody the assertions of McAfee and Brynjolfsson (2012) as they describe increasing value as business shifts towards applying data as a strategy.

Table 1: Three tiers of value creation as a result of Big Data

Tier	Category	Description
1	Data as a Tool	Allows for traditional or core functions and value chains to be improved. Outputs include improved efficiency and effectiveness of existing capabilities and processes. This is where a majority of organisations will find themselves as it is an operational view that is relatively easy to achieve with improvements through real-time or customized decision making and individualised consumer focuses.
2	Data as an Industry	As a majority of firms do not possess the in house capability to appropriately access and leverage their data, many in-source specialised entities focussed on acquisition, storage, construction of infrastructure, processing and analysis of data, and development of software devoted to managing Big Data.
3	Data as a Strategy	Few organisations are dedicated to building data resources which can direct development of radical or innovative business models that link traditional and modern strategic thought. This relied on business leaders being evidence based or data led in their decision making and strategy development. They focus on data flow rather than data stocks to create leverage through insight that allows them to leave competitive barriers.

Modified from Mazzei and Noble (2017)

The first tier applies Big Data as a tool to improve and facilitate efficiency in existing capabilities within the business (Mazzei & Noble, 2017). The vast majority of the examples of value adding applications presented above fall within this tier due largely to the combined factors of; firstly organisational leaders seeing data and analytics as a resource which is positioned to resolve or improve existing paradigms in a given business it being a far simpler proposition. Secondly, application achieving applications within the subsequent tiers requires appropriate resources, scarce skills, business alignment and not only the will but the ability to redirect business activities based on insights from data (Akter *et al.*, 2016; Boncea, Petre, Smada, & Zamfiroiu Anagrama, 2017; Comuzzi & Patel, 2016; Lismont *et al.*, 2017; Mazzei & Noble, 2017).

The second tier describes Big Data as a stimulus for industry in and of itself. As Big Data capability is predicated on highly specialised skill sets and systems infrastructure and computational ability, it enables establishment of entities which focus on infrastructure management and construction, acquisition of data, storage, development of software, analysis and associated support services (Mazzei & Noble, 2017). Companies that have begun to

operate in this space include Amazon, Cloudera, Hortonworks, Capgemini, Microsoft, Palantir Technologies, Pivotal and several of the traditional consulting type organisations.

The final tier is one that houses comparatively fewer organisations, but is also the sphere where greatest potential for the application of Big Data resides. These are companies which not only allow data to influence their operations, but which actively develop ecosystems which allow for knowledge and resources to be directed and mobilised as a result of access to data. Mazzei and Noble (2017) cite companies such as Apple, Alphabet (Google), Amazon and Facebook as examples who have been able to divorce themselves from many competitive influences to their products and services as a result of being devoted to actions based on data.

Though the vast majority of businesses confine themselves to actions within tier 1, there are many examples where data can become a strategy driver for business, as was described previously. Achieving this is often not a simple proposition however and appears to be a source of much of the poor conversion of potential to realised value in literature and business.

2.2.7. Common challenges in the Big Data environment

Application of data driven decision making directly improves daily business function (Brynjolfsson *et al.*, 2011; Lavallo *et al.*, 2011; McAfee & Brynjolfsson, 2012; Nannetti, 2012), which in turn means that data, and analytics should be viewed as a strategic capability, central to the functioning of a business. Achieving this appears to be a significant challenge for a large number of organisations however, who by inference, rely on intuition-based rather than data-based decision making. On average only half a percent of the Big Data that organisations have available at their disposal is being analysed for insight (Bradley *et al.*, 2013) and up to half of all Big Data projects in large organisations are never completed (Comuzzi & Patel, 2016).

If considered broadly, the predominant challenges reported in the Big Data environment can be classified as being either technical or non-technical in nature (Akter *et al.*, 2016; Chen *et al.*, 2014; Dutta & Bose, 2015; Frizzo-Barker *et al.*, 2016; Jin, Wah, Cheng, & Wang, 2015; Loebbecke & Picot, 2015; Wamba *et al.*, 2017, 2015; Wang *et al.*, 2016). *Technical* challenges described by those authors include system infrastructure and architecture, data storage, data curation, data complexity, security, compliance, analytical tools and techniques, technical resources, computational and processing complexity, visualisation, scalability to name but a few. *Non-technical* challenges include; access to data, skills, decision making, resource management, executive support, understanding of outputs and capabilities by business people, business structures, strategic alignment, regulation, how to assist analysts with

complex issues, stakeholder buy-in, implementation of systems and capabilities and change management.

Frizzo-Barker *et al.* (2016) presented a detailed empirical study on the state of Big Data research in business literature and found that the three concurrent core areas of concern that have been reported. The first and most prominent of these was a lack of appropriate skill sets and tools required to take advantage of the unprecedented scale of data available such that Big Data deliverables can be achieved. As the bulk of research activity in the Big Data environment is currently focussed on this area as previously stated, it is also the area that is currently receiving the greatest attention, and will not be considered in further detail in this review.

The second most prevalent concern was data privacy and regulation. This area is becoming ever more prevalent given the amount of personally identifiable information being shared in the modern era, the increased awareness of consumers to this, and the active development of related legislation, such as the POPI act in South Africa (Alharthi *et al.*, 2017; Kambatla *et al.*, 2014; Kshetri, 2014). As was stated previously, though of significant concern, this area is not to be considered in detail in this review as describes considerations beyond the unit of analysis of this study, being the firm or organisation, as legislation describes the environment within which the business will function. Furthermore, legislative concerns are a topic which will be more appropriately tackled by researchers trained in this discipline.

The third area highlighted the need to improve business systems and organisational dynamics in order to enable effective execution of Big Data capabilities, a factor also specifically highlighted in a recent paper directed at addressing challenges in Big Data implementation by Alharthi, Krotov, & Bowman (2017).

By nature, Big Data projects are broad reaching as data are not only generated at all levels of an organisation, but the outputs of Big Data programs should be implemented across multiple divisions and decision making levels in an organisation (Chen *et al.*, 2014; Comuzzi & Patel, 2016; Wang *et al.*, 2016). This thus requires cultural appreciation and acceptance of the platform by the organisation (Alharthi *et al.*, 2017) in order to deliver appropriate results and avoid execution in silos. This agrees closely with Lavallo *et al.* (2011) who asserted a leading cause of poor Big Data implementation are organisational dimensions of lack of understanding of how to use analytics to improve the business, lack of focus by management with competing priorities, insufficient business processes for both data and implementation and organisational

culture unreceptive to change. These challenges are not unlike other enterprise wide technology deployments such as EPR systems, which themselves are challenging to govern (Comuzzi & Patel, 2016), and thus they should be treated similarly.

As was described at the beginning of this section, Big Data projects are not being well executed in business, despite the clear causal links to improved business performance wide data driven decision making and strategies (Akter *et al.*, 2016; Lavallo *et al.*, 2011; McAfee & Brynjolfsson, 2012; Monino, 2016; Power, 2015). Though advanced capabilities (e.g. advanced applications of machine learning and artificial intelligence) remain a challenge only achievable by a scare number of skilled personnel with appropriate systems resources, in most cases the base capabilities to enable Big Data are relatively easily attainable. This is particularly true with outsourcing of technical requirements and cloud based storage and computing readily available to the broader market through service providers such as AWS and Azure. Due to this, it is asserted that perhaps a more telling factor limiting execution of relatively straightforward Big Data initiatives is an artefact of organisational issues rather than skills and technological issues which are prominent at the leading edge of the field.

A core purpose of this research is to move towards clarifying and alleviating concerns of limited success in implementation of Big Data capabilities as a result of the characteristics of organisational dynamics. This is proposed through further characterising conditions in the business, particularly regarding perception, acceptance and application of BDA by business people in relation to their business structure. It is proposed that if these are aligned, BDA capabilities will be effectively positioned and will allow for improved delivery of value by virtue of being within an amenable business environment. This does however do so assuming that technical capabilities and appropriate legislative compliance which are required to enable a BDA capability are in place.

As will be described in further detail hereafter, there is a developing body of Big Data organisational literature. McAbee *et al.* (2016) have detailed that, though fundamental to advancing the field, there is an over emphasis on deductive and theoretical literature in this arena. There is thus a requirement to substantiate the causal relationships that have been asserted between application of Big Data environments and organisational maturity. The following sections will consider the theoretical bases of organisational maturity as influences by propensity for business model innovation, decision making theory and dynamic capability theory in accordance with this position.

2.3. Business Maturity

Few organisations have been able to successfully and sustainably leverage Big Data to deliver value across their entire organisation (Comuzzi & Patel, 2016; Elgendy & Elragal, 2016; Lavallo *et al.*, 2011; Mazzei & Noble, 2017). Though technical complications have been described, it is asserted that there is a causal link between the ability of an organisation to achieve this and the extent of that organisations maturity as a measure of the ability to effectively incorporate BDA as a new business capability within a going concern. To this end maturity models are reviewed to inform this position.

2.3.1. Maturity models in the context of this study

Maturity models that are of interest in this study pertain to attributes within organisations which allow for them to adopt and assimilate capabilities effectively, which is reported to be a function of 'maturity'. Though they are not directly applied as models in this study, their concepts applied are of value in describing the capability. These models tend to define maturity along one of three lines. Firstly, the extent to which processes are defined, organised, managed and controlled by an organisation, as applied by Paulk *et al.* (1993). The second is that of the capability of human capital within an organisation and how effectively they proliferate knowledge creation and acceptance as in Dong *et al.* (2011) and Perez-Soltero and Soto (2017). The last is product maturity, as measured by the proficiency and efficiency by which processes are completed and deliverables met or products delivered (Kosieradzka, 2017; Rajterič, 2010). These themes form the basis on which the organisational constructs assessed in this research were selected.

The following themes were developed for this research. In relation to the first bracket defined above (extent to which given processes are defined, organised, managed and controlled), ***firm dynamic capability (FDC) and firm absorptive capacity (FAC)*** are considered to measure how well business is able to implement understand and incorporate new capabilities into the business, and to adapt to changes in the business environment. The second bracket (organisational capability of human capital towards knowledge creation and acceptance) is assessed through ***organisational culture to BDA capability (OCBDA)***, particularly towards assessing how accepting business is to new developments and capabilities. The final bracket (product maturity or proficiency of products delivery), in this case measured by usage of insights generated by BDA capability rather than gut feel, as assessed through ***data driven decision making (DDDM)***.

A major limitation of a majority of the models described is that the variables and rationale against which they are produced are typically not well described, which thus makes research based on these models challenging (Halper & Krishnan, 2014; Hortonworks, 2016; Rajterič, 2010). It is for this reason that the constructs described above are assessed directly, and their implication with regard to whether a BDA capability is being effectively implemented will be discussed against the scaffold of business model theory.

2.3.2. Big Data maturity models

BDA is by in large an amalgam of traditional data analytics, business intelligence, capability management, information and knowledge management. Though many of the same principles apply, given the reach and complexity of the Big Data environment, there are however areas of greater complexity (Boncea *et al.*, 2017; Comuzzi & Patel, 2016; Lismont *et al.*, 2017; McAfee & Brynjolfsson, 2012). Several Big Data and data science maturity models have been developed by business entities in literature, including Hortonworks (2016) Big Data maturity model, Domino Data Lab (n.d.) data science maturity model, Accenture Digital (2015) analytics maturity model, TDWI Big Data maturity model guide (Halper & Krishnan, 2014) as well as Strategy& (formerly Booz and Co) Big Data maturity model (El-Darwiche, Koch, Meer, & Tohme, 2014). Though good for direct benchmarking by a business, these software vendors or consultancy firms appear to apply, or at least report less academic rigour than is required for an academic application, and may well have been produced for business as a marketing exercise (Rajterič, 2010).

Though Crowston and Qin (2011) produced a capability maturity model for scientific data management only recently has directed Big Data maturity model literature been developed. These includes Lismont *et al.* (2017) who characterised Big Data maturity indicators, Comuzzi and Patel (2016) who provide a well-constructed Big Data maturity model and Boncea *et al.* (2017) who provide a maturity analysis for Big Data technologies.

As Boncea *et al.* (2017) report on maturity of Big Data technology, there is limited consideration of the broader organisational environment in which the Big Data platform is placed. This study focuses on assessing how BDA is being applied within organisations with respect to aspects of organisational decision making, the processes (business capability), capabilities and application (people) and culture (acceptance of new capabilities and innovation in relation to business as usual). The Model proposed by Comuzzi and Patel (2016) provides a relatively good base in this regard. As is noted by them, though the levels of

maturity applied are considered appropriate, there may be more factors required to appropriately account for an organisations Big Data maturity. Under this auspice, these parameters are to be specifically assessed in this research, and discussed in further detail hereafter. These factors lend further credence to the position which this research takes.

2.4. Business Operating Models

Given that the focus area of this research is the implementation of Big Data platforms with respect to a broader organisational context, the following section is included to both describe the broader organisation context in which this will play out, and the environment which will need to be altered to appropriately implement a Big Data platform.

2.4.1. Description of the business model concept

Though the concept of a business model is cited as early as 1957 (Bellman, Clark, Malcolm, Craft, & Ricciardi, 1957), it is a topic of ongoing development. The major development from initial descriptions is that the term ‘business model’ has shifted from describing an operational or process driven environment to describing a strategic management tool that allows for dynamic developments and innovation to be incorporated into a business (Baines, 2014; Baldassarre, Calabretta, Bocken, & Jaskiewicz, 2017; Eurich, Weiblen, & Breitenmoser, 2014; Frankenberger, Weiblen, & Gassmann, 2014; Gassmann, Frankenberger, & Csik, 2013; Ghosh, 2016; Štefan & Branislav, 2016). BDA is considered such a development or innovation within this context, and thus is a strategic tool which must be appropriately applied by management within a business environment.

As with Big Data, this is a developing field which has seen significant interest and associated development in literature in recent years. As a result of this there appears to be no generally accepted definition. What is common to all positions is the relevance of a business model to corporate strategy and firm performance, particularly with regard to incorporating innovation and maintaining competitive advantage, and furthermore that it is a conceptual rather than operational description of the components of the organisation (Bereznoi, 2014; Eurich *et al.*, 2014; Gassmann *et al.*, 2013; Štefan & Branislav, 2016; Teece, 2010).

Closely related, but different to operating models, business maturity describes an ability of business to appreciate, accept and incorporate new capabilities into the business. These thus are closely aligned and important concepts with regard to this study, but differ significantly in application in this context.

2.4.2. Business model development and innovation

A large portion of recent business model literature centres around business model innovation as a way to create and capture value (Baldassarre *et al.*, 2017; Eurich *et al.*, 2014; Frankenberger *et al.*, 2014). As Teece (2010) describes, the business model embodies the businesses architecture or structure into which innovations must be appropriately nested. Given the pace of innovation and development in today's world, organisations must be flexible enough to update and improve their business models in order to stay competitive (Teece & Leih, 2016). This study applies the approach that the business model described the environment into which a capability (like BDA) is inserted.

As with any appropriately significant innovation, it will only become a commercial success if it is embodied within an appropriately configured architecture to allow it to disrupt market (Bereznoi, 2014; Teece, 2010). Technological developments, such as Big Data and BDA, are often described as both innovations in their own respect, and as drivers of innovation. In order for them to achieve desired success as well as sustained impact within a business, they should thus be viewed as component of a business's architecture and must be appropriately configured into business models (Bereznoi, 2014; Teece, 2010). Given its impact, businesses should not just engage in business model innovation, but should rather embrace it as a core competency to facilitate innovations to deliver competitive advantage (Wrigley, Bucolo, & Straker, 2016).

2.5. Inhibitors to Implementation of Innovation in Business

As business model innovation must persist for businesses to remain relevant and competitive, it is an iterative process which results in business models never being complete or static (Bereznoi, 2014; Teece, 2010; Wrigley *et al.*, 2016). Given the strains that this state of flux can have on a business, there are several areas which may impede it, and which are relevant to implementation of Big Data platforms in business models including;

- **Implementation complexity;** particularly with systems that not only complex in their own function and requirements, but which span across all divisions and levels of business such as Big Data does (Comuzzi & Patel, 2016).
- **Inadequate leadership and backing** of project or initiatives, particularly by the executive level, and in particular the CEO of an organisation has been cited regularly as an inhibitor of Big Data platforms in organisations The literature also highlighted the issue of inadequate leadership as a barrier to innovating an organisation's business model (Akter *et al.*, 2016;

Colas *et al.*, 2014; Lavallo *et al.*, 2011). This is described both from a need for clear goal setting and management of individuals toward attaining that, and in terms of initial buy-in and sustained drive to roll-out the initiative (Colas *et al.*, 2014). As is described by Von Den Driesch, Eva, Flatten, and Brettel (2015), the influence of social capital on dynamic capabilities a business is significant, and is definitive at the CEO level.

- **Deficiencies in organisational culture;** Leidner and Kayworth (2006) assessed the influence of culture on adoption of IT technology at national, organisational and group levels. They found that unless organisational culture is aligned with the embedded traits of the system, it inevitably leads to conflict. Choosing and directing what and how to align requires strong leadership and alignment to strategic goals within the organisations environment. It also requires structured assessment of attitudes to, and engagement with the proposed organisational change (Shah *et al.*, 2017). As is described by Loebbecke and Picot (2015), given the widespread potential of digitisation, robotics and Big Data analysis, there is an ever increasing impact of the sector on routine functions in business, which will engender uncertainty and concerns for many within an organisation, particularly if it impacts on their job security. A further area where this is evident is in the face of prevailing dominant logic, where managers conform to rules, norms and beliefs to guide their action (Wrigley *et al.*, 2016). This logic actively discourages innovation and may result in critical opportunities being missed and competitive advantage being lost over time.
- **Lack of organisational structures and processes** are often reported to hinder the adoption and capability to adopt innovations such as Big Data (Chen *et al.*, 2014; Colas *et al.*, 2014; Lavallo *et al.*, 2011; Teece, 2010). It goes without saying that unless a business model and organisational architecture are structured in a way that allows for an innovation or capability to be implemented, then the ability of a firm to be dynamic is impeded severely.
- **Poor strategic alignment** refers to a mismatch between the strategic positioning of an organisation and the design of the activities that support this organizational strategy (Roelens, Steenacker, & Poels, 2017), in essence describing that the innovation does not directly support outputs of the existing strategy.
- **A lack of strategic flexibility** of a firm refers to a limitation in a firm's ability to quickly and sustainably adapt to dynamic changes in market (Schneider & Spieth, 2014). Therefore, a critical enabler for a firm in ensuring its ability to transform itself continuously is the ability to remain dynamic, to keep options open and to implement changes with minimal organisational upheaval (Mazzei & Noble, 2017; Teece, 2010; Teece, Peteraf, & Leih, 2016).

2.6. Assimilation of Innovations and Capabilities into Organisations

The process of adoption or assimilation of innovations and capabilities within an organisation is a well researched aspect of business literature, particularly in the IT domain, with significant developments since its inception in the early 1980's (Elbashir, Collier, & Sutton, 2011; Rad, Nilashi, & Dahlan, 2017; Wisdom, Chor, Hoagwood, & Horwitz, 2014). In today's business environment, characterised by globalisation, rapid technological developments and the emergence of new economies, the ability to effectively adopt and assimilate innovations is essential to retain market competitiveness (Ben-Menahem, Kwee, Volberda, & Van Den Bosch, 2013; Roberts, Galluch, Dinger, & Grover, 2012). Researchers are, in some cases, even calling for the need for radical innovations rather than incremental innovations to allow firms to sustain their long-term competitive advantage (Agostini, Nosella, & Filippini, 2017).

Research has focussed on adoption at both the individual and organisational level, with a number of frameworks and conceptual models resulting from consideration of numerous dependant and independent variables (Lai, 2017; Rad *et al.*, 2017; Wisdom *et al.*, 2014). Given the sustained development of IT related technologies and the prominence of their impact on both business and society, the study of adoption and assimilation of technologies has received significant increase in research interest and publications since the mid 2000's (Rad *et al.*, 2017). BDA is one such capability which is a subset of this field, but which itself has received limited attention in academic literature (Côrte-Real, Oliveira, & Ruivo, 2016; Oliveira, Thomas, & Espadanal, 2014). Furthermore, as described by Agostini *et al.* (2017), though external (to a given business) sources of innovation (like a big data capability) are typically the focal point of research, there is often a gap between the external innovation and businesses internal organisational context. Addressing this gap in business is a fundamental requirement that required research to better appreciate how to implement innovations, particularly if they are radical in nature.

Assimilation of capabilities at an organisational level, though complex, has been characterised as a stepwise process by several authors. The processes are however typically segmented into phases. The initial phases include recognition of a need by decision makers, followed by a searching for solutions, then a decision to adoption of the solution (Rad *et al.*, 2017; Wisdom *et al.*, 2014). The second phase includes the actual use or application of the innovation, and then the continuance of use (Rad *et al.*, 2017; Wisdom *et al.*, 2014).

The initial adoption process has been further characterised as including the processes of pre-adoption (awareness of innovation), peri-adoption (continued access to detail of the innovation), and established adoption (commitment to the adoption decision) (Greenhalgh, Robert, Macfarlane, Bate, & Kyriakidou, 2004). Following initial adoption, the process moves to initial implementation activities (Frambach & Schillewaert, 2002; Wisdom *et al.*, 2014). Though little research is available on rejection, implementation involves the decision to pursue the adoption, as well as the staff's acceptance of this, followed by initiation of their individual processes of accepting and applying the innovation (Frambach & Schillewaert, 2002). Both phases are often punctuated by shocks setbacks and surprises, which if too large, may result in rejection of the innovation, though there is limited literature on that process (Frambach & Schillewaert, 2002; Greenhalgh *et al.*, 2004; Wisdom *et al.*, 2014).

Among the most commonly cited challenges toward effective implementation of BDA include the softer aspects of organisational culture (understanding, acceptance or impetus to use a new capability or innovation), and inflexible organisational designs and strategic outlooks that block implementation (Comuzzi & Patel, 2016; Lavallo *et al.*, 2011; Loebbecke & Picot, 2015; Mazzei & Noble, 2017; McAfee & Brynjolfsson, 2012). Toward this, the three aspects focussed on below describe aspects of the initial phase of adoption. These include organisational culture toward innovation (acceptance of capability) followed by dynamic and absorptive capability theories, which describe the ability and flexibility of business to incorporate new capabilities into their structure. The following section considers conditions toward the second phase of adoption, being application or implementation of innovation, considered through organisational knowledge management and organisational decision making theories.

2.6.1. Organisational culture toward new capabilities and innovation

Organizational culture, as expressed through traits such as norms, values, and dispositions are central to the way in which businesses function on a daily basis, and by association, on how these traits may be changed (Leidner & Kayworth, 2006). The extent of the impact of culture on business operations cannot be underestimated, even when it is blatantly obvious that the norm may be counterproductive. This may be illustrated by there being a culture to refrain from exposing medical errors among medical professionals, despite errors being responsible for 11% of death in Australia and more deaths than either HIV or motor vehicles in the USA (Leidner & Kayworth, 2006). On the positive side, an appropriate culture is widely attributed as being a core pillar of organisational success (Donate & Guadamillas, 2010; Guerra & Knox, 2008; Leidner & Kayworth, 2006; Martín-de Castro, Delgado-Verde, Navas-López, & Cruz-

González, 2013; Rogers & Meehan, 2007; Shahzad, Xiu, & Shahbaz, 2017). This has been illustrated by Dell's founder, Michael Dell and CEO Kevin Rollins who, along with their business model, strategy, talented staff and operational discipline, attribute their organisational culture as being the major difference in their business (Rogers & Meehan, 2007).

As was described by Martín-de Castro *et al.* (2013), culture may be defined as a “*deeper level of basic values, assumptions and beliefs that are shared by the members of an organisation*”. They furthermore defined an innovation culture as being “*the shared common values, beliefs and assumptions of organisational members that could facilitate the product innovation [and innovation adoption] processes*”. Though this study focuses on the organisational level, there is also a significant impact of culture and characteristics of the individual which must also be considered for both phases of adoption described previously (Crossan & Apaydin, 2010; Greenhalgh *et al.*, 2004; Leidner & Kayworth, 2006; Wisdom *et al.*, 2014). Within the first phase (pre-adoption) the theoretical frameworks of shared professional values and customer-centeredness are highly correlated (Wisdom *et al.*, 2014). Within the second phase (adoption), these same two frameworks, as well as additional frameworks of a culture of problem-solving are found to be positively correlated (Wisdom *et al.*, 2014).

In the ever changing dynamics of today's business environment, it appears abundantly clear that having a these aspects of organisational culture entrenched is crucial so that they may enable adoption or assimilation of new capabilities and innovations. The readiness of an organisation for change (through the presence of these cultural norms) is considered a crucial precursor to the successful implementation of complex changes, such as a Big Data capability (Weiner, 2009; Wisdom *et al.*, 2014). Indeed, a failure to establish readiness may account for half of all unsuccessful, large-scale organizational change efforts (Weiner, 2009).

Organisational characteristics represent the intersection of the business environment and staff of the business, and accordingly is an area often described by contradictory findings (Damanpour, 1991; Leidner & Kayworth, 2006; Wisdom *et al.*, 2014). Factors such as leadership support for an innovation as leadership's experience with adoption leads, tends to lead to better adoption, however hierarchical leadership may hinder adoption (Crossan & Apaydin, 2010; Elbashir *et al.*, 2011; Wisdom *et al.*, 2014). Choosing and directing what innovations to adopt and how to align business to do so requires strong leadership, as well as alignment of capability to strategic goals (Elbashir *et al.*, 2011; Shah *et al.*, 2017). Furthermore, a structured approach to assessment of attitudinal and behavioural engagement in relation to

proposed organisational changes (Shah *et al.*, 2017). Leaders create cultures conducive both to internal innovation and acceptance of external innovations through shared vision which is clearly stated, attainable and valuable, promoting autonomy of staff, enabling calculated risk taking, and driving motivation (Agostini *et al.*, 2017; Crossan & Apaydin, 2010; Wisdom *et al.*, 2014). These traits positively impact on the relationship between open innovation and innovation performance by fostering an environment for employees to take ownership of developments and to take a variety of decisions without a strong supervision and control (Agostini *et al.*, 2017). Further factors which may be contradictory include that the presence of research infrastructure and additional resources tend to facilitate adoption, but if the organisational structure is too rigid, or requires excessive input from individuals, likelihood of success of adoption decreased (Leidner & Kayworth, 2006; Wisdom *et al.*, 2014).

Traits of innovations themselves that influence employees acceptance toward adoption include ease of use, improvements to existing practices, that it is observable, cost-effective, low risk, adaptable to the needs and norms of users and the organisation, it is evidence-based, and relevant (Donate & Guadamillas, 2010; Frambach & Schillewaert, 2002; Greenhalgh *et al.*, 2004; Wisdom *et al.*, 2014). These traits describe characteristics of direct perception by employees. The underlying culture of the organisation is a significant factor towards employees' attitudes and motivations towards adoption (Wisdom *et al.*, 2014). Though communication, engagement and feedback regarding innovations and the adoption process have been shown to increase adoption significantly, the characteristics of individuals are also significant factors (Greenhalgh *et al.*, 2004; Wisdom *et al.*, 2014). Pre-adoption includes traits of awareness of innovations (supported by a culture of looking for them), self-perception of organisational attributes and needs, innovativeness, tolerance of ambiguity, propensity to take risks, available skills and experience (Agostini *et al.*, 2017; Crossan & Apaydin, 2010; Shahzad *et al.*, 2017; Wisdom *et al.*, 2014). Traits within the adoption phase continue these characteristics, but include individual knowledge base, experience, knowledge sharing and networking and propensity for risk-taking (Agostini *et al.*, 2017; Wisdom *et al.*, 2014).

Cognisance of these traits and their influence on culture and the propensity for absorption are increasingly driving firms to focus on these attributes and actively improve communication, networking and exposure to capabilities which are intended to drive further innovation and adoption (Agostini *et al.*, 2017). The ability to actively do so does however suggest strong measure of maturity of the business, in process maturity, people maturity, object maturity or a combination thereof (Mettler, Winter, Mettler, Rohner, & Winter, 2010).

Leidner and Kayworth (2006) assessed the influence of culture in relation to adoption of technology at national, organisational and group levels and the various conflicts that are encountered therein with the IT environment. Of significance therein is the finding that unless organisational culture is aligned with the embedded traits of the innovation, problems inevitably arise. As is described by Loebbecke and Picot (2015), given the prominence and potential of digitisation, robotics and Big Data analysis, there is an ever increasing impact of the sector on routine functions in business. This illustrates the desire for such capabilities to be incorporated into businesses gradually shifting to a requirement. Unless cultures are aligned, or BDA platforms are malleable enough, this will likely engender uncertainty and concerns. Furthermore, where organisational culture is that managers conform to pre-existing rules, norms and beliefs to guide their action, absorption of innovation may be hindered (Wrigley *et al.*, 2016).

2.6.2. Absorptive Capacity of Organisations

Since its initial appearance in Cohen and Levinthal's (1990) seminal work, absorptive capacity appears to be defined, with reasonable consensus, as being a function of the firm's prior related and complementary knowledge. Further description on the relates to a firm's ability to *identify existing external knowledge that is useful (pre-adoption), incorporate this into the firm and apply it through innovation and competitive actions (adoption)* (Cohen & Levinthal, 1990; Elbashir *et al.*, 2011; Gao, Yeoh, Wong, & Scheepers, 2017; Greenhalgh *et al.*, 2004; Roberts *et al.*, 2012; Wisdom *et al.*, 2014). Furthermore, organisations with good pre-existing knowledge and skills (mature business capabilities) are more likely to incorporate further knowledge or innovations (Wisdom *et al.*, 2014).

As described above, assimilation of new capabilities or innovations into business is heavily dependent on the employees and leaders of that business. Given this, research of adoption needs to specifically consider how cultural controls influence implementation (Elbashir *et al.*, 2011). Elbashir *et al.* (2011) further define absorptive capacity as being a result of culture. They define value (cultural) controls as the *"explicit set of organizational definitions that senior managers communicate formally and reinforce systematically to provide basic values, purpose, and direction"*. They submit that the resulting control that leadership driven culture imparts on knowledge management and resource development governs the *"identification, absorption, and strategic application of new external information"*, which is absorptive capacity. Though this position agrees with the initial assertion by Cohen and Levinthal (1990) insofar that the impact of absorptive capacity on a firm is by individual members is significant regarding

assimilation of new external information, they do not attribute it solely to leadership positions. Rather, Cohen and Levinthal (1990) cite research in the behavioural and cognitive sciences regarding learning and knowledge acquisition that describes collective organisational culture and knowledge of individuals.

Regardless of the source of absorptive capacity within a firm, there is consensus that it is a crucial capability that requires significant attention be paid to it, as it is a trait which enables a firm's long term survival, by strengthening, complementing, or refocussing a firm's knowledge base and by increasing their ability to anticipate and appropriately position themselves regarding trends in market (Elbashir *et al.*, 2011; Roberts *et al.*, 2012). Both the positions of Cohen and Levinthal (1990) and of Elbashir *et al.* (2011) clearly suggest that absorptive capacity can be influenced and improved by investing in the absorptive capacity of individuals through by influencing culture, improving networking and exposure to new capabilities and investing in research and design as a capability (Gao *et al.*, 2017; Roberts *et al.*, 2012; Zahra & George, 2002). As a result, the assimilation component of absorptive capacity refers to the firm's ability to "*transform, re-configure, and re-deploy resources in order to be ready to exploit new external information*" (Gao *et al.*, 2017).

Through its research and development, firms develop their collective knowledge in areas of market, key technologies and how those areas relate to their business or market (Gao *et al.*, 2017; Roberts *et al.*, 2012). Absorptive capacity has been illustrated to have wide ranging application or impact in business. Examples include application in strategic management (Lane & Lubatkin, 1998), strategic renewal (Ben-Menahem *et al.*, 2013), business process improvement (Manfreda, Kovacic, Štemberger, & Trkman, 2014), reconfigure existing capabilities (Pavlou & El Sawy, 2006), forecasting technological trends (Cohen & Levinthal, 1994; Roberts *et al.*, 2012), creation and acquisition of technology (Keller, 1996), assimilation of enterprise information systems (Elbashir *et al.*, 2011), replenishment of the knowledge base (Van den Bosch, Volberda, & de Boer, 1999), technology transfer in international business (Glass & Saggi, 1998), IT innovation (Fichman & Kemerer, 1997) and IT business value (Bhatt & Grover, 2005).

Absorptive capacity is particularly important to research which relate to technical capabilities such as BI, IT and information systems, and by inference Big Data capabilities, due to the rapid convergence and diffusion of communications, computing, and content technologies (Fichman, 2001; Leidner & Kayworth, 2006; Oliveira & Martins, 2011; Roberts *et al.*, 2012). By recognising

these technical capabilities (systems architecture, DevOps, enterprise systems, data storage and warehousing, open source technologies and infrastructures and processing) as strategic resources, IT investments are being combined with complementary assets (knowledge and personnel resources) to enable digital capabilities that improve the organisations absorptive capacity (Oliveira & Martins, 2011; Ravichandran & Lertwongsatien, 2005; Roberts *et al.*, 2012; Sher & Lee, 2004). Importantly, through developing these capabilities, ‘learning-by-doing’ is enabled which further increases absorptive capacity in relation to the constant stream of development in the field (Roberts *et al.*, 2012).

Given that absorptive capacity is strongly influenced by culture as a factor in and of its self, describes both organisational and individual traits as culture does, and is influenced by similar factors to culture, the two are closely inter-related characteristics. Though individual traits are of impact, they are considered as influencers of organisational culture and absorptive capacity respectively in this study as the study considers organisations as the unit of measure. Both measures are however clearly influential on assimilation of capabilities within business as they are depended to the development of appropriate knowledge, skill and norms (Cohen & Levinthal, 1990; Elbashir *et al.*, 2011; Fichman & Kemerer, 1997; Gao *et al.*, 2017; Leidner & Kayworth, 2006; Roberts *et al.*, 2012). As the development of appropriate knowledge and skills within an organisation is dependent on cultural norms this can be used as a control by the organisation to integrate absorptive capabilities into the organisation’s routine practices (Cohen & Levinthal, 1994; Elbashir *et al.*, 2011; Van den Bosch *et al.*, 1999). In this respect, adsorptive capacity can be a learned or instilled trait that may be considered a competence of the business. Significantly, as described by both Cohen and Levinthal (1990) and of Elbashir *et al.* (2011), these are traits described for individuals, however given the influence of culture, it appears upper management and leadership will play a more influential role.

As is described by Elbashir *et al.* (2011), the sophistication and diversity of the organisations IT infrastructure will impact on assimilation as influenced both by the prevalence of engagement with technologies and because it actively described addition to knowledge base, both of which have been shown to improve absorptive capability (Bhatt & Grover, 2005; Keller, 1996; Manfreda *et al.*, 2014; Van den Bosch *et al.*, 1999). Given this trait, adoption and assimilation of technically complex capabilities such as Big Data capabilities and the outputs of BDA is likely to be influenced by the level of existing knowledge, capability and technical capability in a business, whether directly associated with analytics or not. This is further supported by Elbashir *et al.* (2011), as they describe that operational users of such capabilities (described in

relation to adoption of BI in their case, which is not entirely dissimilar to BDA as far as outputs are concerned) within business directly impact the manner in which systems are deployed, in relation to both information available, and drive to adopt.

2.6.3. Dynamic Capabilities of Organisations

Although research and literature toward the concept of Dynamic Capability Theory (DCT) predates the seminal articles by Teece, Rumelt, Dosi, and Winter (1994) and later Teece, Pisano, and Shuen (1997), it was not until these publications that the dynamic capabilities view began to generated substantial interest and a growing flow of research. Since then, the past decade has seen DCT become a highly effective lens for both strategic management and IT management fields (Côrte-Real *et al.*, 2016; Mayer & Teece, 2008; Schilke, 2014; Teece & Leih, 2016). As a field still in its relative infancy it, much like Big Data, it is one that has not yet reached consensus in terms of a definition (Helfat & Peteraf, 2009). This is not surprising given that, according to Barreto (2010) over 1500 articles were published on the topic in the 10 years following the work by (Teece *et al.*, 1994) and (Teece *et al.*, 1997). There is furthermore contention as how it should be viewed. Teece *et al.* (1997) view dynamic capabilities as being “*idiosyncratic factors which give rise to an organisations sustainable competitive advantage*”, and thus have the view that outputs should be viewed as firm performance. On the other hand, Eisenhardt and Martin (2000) view dynamic capabilities as “*best practice*” or “*commonalities between firms*”, and hence view it as a process which give rise to sustainable competitive advantage through the resource configurations that they create.

For the purposes of this report, the view taken by Teece *et al.* (1997) is most appropriate as it suggests enablement of outputs of Big Data platforms directly impact firm performance and competitiveness, as has previously been asserted. Furthermore, this definition more appropriately describes the fundamental capability described by DCT, which describes the ability of a firm to react to or address major environmental changes (Barreto, 2010; Teece & Leih, 2016). More specifically, this refers to the ability of the business to make alterations to its own internal operational, technical, knowledge or strategic capabilities either pre-emptively or in response to shifts in conditions in the business environment (Barreto, 2010; Eisenhardt & Martin, 2000; Helfat & Peteraf, 2009; Sher & Lee, 2004; Tavallaei, Shokohyar, Moosavi, & Sarfi, 2015; Teece *et al.*, 1997; C. L. Wang & Ahmed, 2007; Zahra & George, 2002). Failure to do so describes an inability to adsorb or assimilate capabilities or innovations, which will usually result in a negative impact on the firm’s performance and competitive ability (Barreto, 2010). As described by Teece *et al.* (1997), dynamic capabilities provide an explanatory tool to

analyse a firm's ability to build the competitive advantage in the time of uncertainty and change.

DCT is largely rooted within the theory of Resource Based View and Knowledge Based View, and argues that the idiosyncratic factors held by an organisation allow them to alter their activity in relation to market factors to maintain their competitive advantage (Eisenhardt & Martin, 2000; Helfat & Peteraf, 2009; Teece *et al.*, 1997). Despite the debates around definitions and views, there is conformity on the primary elements that constitute the dynamic capabilities framework which were initially described by Teece *et al.* (1997). These include;

- **Sensing:** the ability to view, identify and interpret opportunities and threats in a business environment (Mikalef & Pateli, 2017; Teece *et al.*, 1997).
- **Coordinating:** the ability to manage and ideally synchronise resources, stakeholders, deliverables and tasks in relation to tasks or requirements. Through this capability opportunities may be seized and synergies identified both inside and outside of the organisation which can be capitalised on further through collaboration (Mikalef & Pateli, 2017; Teece *et al.*, 1997).
- **Learning:** The capacity to gather, understand and exploit knowledge such that improved decisions may be made (Mikalef & Pateli, 2017; Pisano & Teece, 2007; Teece *et al.*, 1994)
- **Integrating:** the process that allows for more efficient problem resolution by more effectively combining the various resources of the organisation (Mikalef & Pateli, 2017; Teece *et al.*, 1997).
- **Reconfiguring:** The ability of the organisation to make strategic decisions and rapidly enact or execute against these by repositioning resources to better align the organisation with the external or market environment (Eisenhardt & Martin, 2000; Mikalef & Pateli, 2017; Teece *et al.*, 1997).

Though the body of literature concerning DCT has grown substantially, with significant areas of influence as described above, empirical research concerning how dynamic capabilities positively influence firm performance appears relatively limited (Giniuniene & Jurksiene, 2015). Reported or deficiencies within DCT are reported, including the failure to incorporate dynamic capabilities into the internal processes within the firm (Drnevich & Kriaucunas, 2011). Given the range of attributes which it covers, complexity of the application of dynamic capabilities and the potential need for extensive management has also been cited (Drnevich &

Kriaucunas, 2011). Literature has also tended to consider dynamic capabilities as having a central role pertaining to the change of internal conditions or characteristics of the firm, whilst the locus of change has varied across elements (Giniuniene & Jurksiene, 2015). These have include resources and capabilities (Eisenhardt & Martin, 2000; Teece *et al.*, 1997), resources and routines (Helfat & Peteraf, 2009). Giniuniene and Jurksiene (2015) furthermore describe that some researchers describe a two-level hierarchy of “zero-level” capabilities (ordinary capabilities that allow a firm to run in the short term, and substantive capabilities which are used to solve a problem) and “higher-level” capabilities, such as dynamic capabilities (capabilities that operate to change ordinary capabilities).

There is a link required between the presence of rapidly changing external environments and the existence of dynamic capabilities (Teece *et al.*, 1997). This was refined in later work to advocate for the special relevance to environments that include international or global markets are well developed, where technical change is systemic, where technological and managerial knowledge are poorly developed, and where regulatory or institutional shocks occur (Barreto, 2010; Teece, 2007). Some researchers are however divided about the kind of external environments that are relevant for dynamic capabilities (Helfat & Peteraf, 2009). Literature describes a spectrum of researchers from those who unequivocally ascribe to there being highly dynamic environments, to those who acknowledge the relevance of DCT in both stable and dynamic environments, and those who simply ignore the characteristics of the environment (Helfat & Peteraf, 2009). As described previously, this study applied the base position that the global business environment is becoming progressively more challenging, and requiring ever more rapid, efficient and effective management and assimilation of innovations, knowledge and capabilities as a result of external environments being highly dynamic (Barreto, 2010). This is reportedly a factor of hypercompetitive (D’Aven, 1994) or high-velocity (Bourgeois & Eisenhardt, 1988) environments, which tend to be associated with increasingly frequent shifts in competitive, technological, social, and regulatory domains. This is a position particularly emphasised within technology environments, and is well characterised within the Big Data environment, which is the interface of multiple rapidly evolving and cutting edge fields. In addition to this highly dynamic environments, Eisenhardt and Martin (2000) argued that dynamic capabilities are also relevant in moderately dynamic markets where change occurs along predictable and paths. In so doing, this explains a capability which ascribes to adoption of a capability within a multitude of industries.

In essence, the major faults of presented with dynamic capabilities research surround a certain lack of clarity, oversimplification of dynamics, under studied measurement of issues and limited empirical support (Helfat & Peteraf, 2009). These so-called 'deficiencies' are the however all merely typical signs of the early-stage development of a theory. Particularly for a theory with such a broad reach, which has not yet been able to fully mature, and in which work largely remains at a foundational level across issues, including definition of the term (Helfat & Peteraf, 2009).

Though DCT may still only provide sufficient basis to be classified as an 'approach' to understanding change and adsorption in business rather than a fully-fledged 'theory' there are clearly identifiable traits of strong theoretical foundations in literature (Helfat & Peteraf, 2009). Of significance in the context of this study is that dynamic capability incorporates a number of associated aspects which influence adsorption of innovation or capabilities within business. Of particular interest is that it incorporates aspects of the softer side of management which are typically underweight Big Data research, which tend to focus on technical attributes rather than the broader business environment and the human aspects of business (Aker *et al.*, 2016; Comuzzi & Patel, 2016; Côte-Real *et al.*, 2016; Frizzo-Barker *et al.*, 2016; Wamba *et al.*, 2017).

As both dynamic capability and absorptive capacity incorporate aspects of organisational human capital, and the resulting impact that they have on business models, these aspects can thus be taken into consideration through its application in a Big Data context (Gao *et al.*, 2017; Roberts *et al.*, 2012; Teece, 2010; Wisdom *et al.*, 2014). In so doing, this theory provides a conduit which may link the outputs of Big Data to the business environment by relating their impact on business model and on innovation thereof. Zahra and George (2002) proposed that absorptive capacity should be reconceptualised as a dynamic capability which pertains to a firm's ability to both create and use knowledge to sustain competitive advantage. Giniuniene and Jurksiene (2015) furthermore describe how dynamic capabilities, innovation and organizational learning are interrelated. This is key as these all feature strongly within organisational culture toward innovation and absorptive capability of the organisation. Furthermore, as an impact, all three are regularly described as having significant influences on business practice, innovation and resulting firm performance, all of which are actively assessed through this study. Dynamic capability thus provides a relatively broadly encompassing perspective towards adoption, and considers more of a link between strategic choices made by a firm in relation to their environment (Barreto, 2010). In this manner it is applied as a

combined term describing a business's ability to take action in response to innovation and absorb that capability, which should improve fitness or competitiveness of the business.

2.7. Assimilation of Innovations and Capabilities in Organisations

Beyond a position of recognising a need to alter the base capabilities of business given shifts in its environment, having the culture willing to accept that change (pre-adoption), and having a business environment capable of absorbing it as part of its typical operations (adoption); success if the capability is typically measured as a factor of whether it is appropriately applied after it is assimilated, and whether this can be sustained by the business (Rad *et al.*, 2017; Wisdom *et al.*, 2014). Within the context of this study, this refers primarily to whether insight extracted from data can be used by business effectively, and whether this can support a sustained business fitness given technological advancements in the business environment, or ideally uplift in firm performance.

Davenport and Patil (2011) argued that one of the most pertinent BDA challenge is the use or application of the data itself by users in business, as the large volumes of 'insight' or 'directives' often overwhelm and confuse the decision maker. This describes a need for decision makers both to be able to receive and focus on the 'right' data, which may be describes as being the data which is reliable, accurate and punctual information relative to the decision that needs to be made (Rajterič, 2010). It is the position of this study that BDA outputs are, at their core, just insights and information from data which must be actioned by business. Essentially, the role of the Big Data capability is to process very large, diverse and in many cases complex sets of data such that relatively direct actions can be derived from them. The proposed availability and application of the 'right' data is thus a key output of BDA, but only on a larger scale than business was typically accustomed as not only is there an immense data set available, but given the issue to be solved, there may be multiple interpretations of the data available. Knowing how to assess the data relative to the issue would be a characteristic of a mature BDA capability within a business, whilst knowing how to apply the BDA outputs would be a characteristic of a mature business management capability. The output of the platform is data or information, however in order to convert the information into knowledge, the information must be understood by managers in context, and appropriate actions initiated. As was previously described, organisations which make decisions based on data are more effective in market than their competition (Lavalley *et al.*, 2011; McAfee &

Brynjolfsson, 2012). The next section considers the theory base of decision making literature with a focus on Big Data applications as inputs to the decision making process.

2.7.1. Organisational Decision Making

From a decision support context, many researchers have asserted that one of the most effective ways for businesses to extract value from Big Data capabilities is through the increased application of BDA as an input to the organisational decision-making processes (Baines, 2014; Davenport & Patil, 2011; Lavallo *et al.*, 2011; McAfee & Brynjolfsson, 2012; Popovič, Hackney, Coelho, & Jaklič, 2012). Among the key means of doing so is that data-based or evidence based decisions making eliminates or negates biases such as those commonly encountered with decisions based on intuition as decision become more consistent, cost effective, replicable, and can be transferred within the business (Davenport & Patil, 2011).

Decision making theory is considered herein as the core output of this capability is in the use of the correct data to drive the best possible outcome for the business. This capability is one which has been present in literature since data began being processed to inform action. Simon's four phases of decision making is typically adopted by decision makers, and includes intelligence, design, choice, and implementation (Elgendy & Elragal, 2016). Though Big Data applications have increased the complexity of the analytical processes themselves, and the business environment will in many cases be more complicated, the fundamental concept of applying data in decision making remains. As is described by Elgendy and Elragal (2016), the multiple steps a complexity introduced with application of Big Data platforms, as well as the increased number of participants in the process does however make appropriate application more challenging.

2.7.2. Big Data and Data Driven Decision Making

The rise of BDA has prompted an evolution in the types of information that is available from data, and thus also influences how data should be utilised in organisational decision making. This is however an aspect that many organisations appear to struggle to achieve (Boyd & Crawford, 2012; Elgendy & Elragal, 2016; Janssen, Voort, & Wahyudi, 2017; McAfee & Brynjolfsson, 2012; Provost & Fawcett, 2013; Rajterič, 2010). Given the previously stated value that is derived from appropriately implementing these capabilities in a business, it is imperative for business leaders today to understand how to apply Big Data in decision making. Furthermore, and as importantly, implementation of dig data requires appropriate decision

making by business managers (Alharthi *et al.*, 2017; Janssen *et al.*, 2017; Provost & Fawcett, 2013; Y. Wang, Kung, & Byrd, 2015; Wang *et al.*, 2016).

Normative decision theory focuses on finding methods and tools to direct the best choice to make, and is based on the assumption that the 'decision maker' is rational (Wang *et al.*, 2016). Informed by this position, each step of the Big Data process from data acquisition through to data visualization is also informed by derived knowledge, thus this is an integral capability within the entire data value chain, and affects persons who interact with the data at each stage of the process.

As is described by Janssen *et al.* (2017), the capabilities of each entity that interacts within the Big Data chain can influence the outcomes of the subsequent decisions which may be made based on that data. these decisions include aspects from what data to acquire, how best to collate, what combinations of data are appropriate for the question posed, how to mitigate for incomplete or divergent data sets, how to clean data and then integrate selected data sets, what ethical considerations may be in the application of data, how to analyse the data and then how best to make decisions in applying the outputs of those processes (Elgendy & Elragal, 2016; Janssen *et al.*, 2017; Monino, 2016; Wang *et al.*, 2016). Managing this thus requires both technical capabilities and skills and processes. Wang *et al.* (2016) provide a comprehensive assessment of technological controls and procedures, which include tools such as statistics, machine learning, data mining and visualisation. As the capabilities of firms to collect, prepare, process, analyse, visualise and apply data may be different, this suggests a level of improved capability correlating to increased maturity of the system (Elgendy & Elragal, 2016; Janssen *et al.*, 2017; Wang *et al.*, 2016).

Considering application of data, the hierarchical model describes the increased value moving from data, to information and finally to knowledge (Monino, 2016; Wang *et al.*, 2016). This is key to the current discussion as companies are deriving significant value from non-physical or intangible assets of data that enable knowledge (Monino, 2016). This final stage of the data chain is of particular interest in this study as it considers the application of knowledge generated by specialists accustomed to the science of data analysis by general business.

Chapter 3: Research Questions

“If I had an hour to solve a problem and my life depended on it, I would spend the first 55 minutes determining the proper question to ask” - Albert Einstein

3.1. Research Overview

Though significant attention has been paid technical attributes of the Big Data environment, comparatively less research has considered the impact of the business environment on the success of Big Data Analytics (BDA) deployments (Akter *et al.*, 2016; Alharthi *et al.*, 2017; Chen & Zhang, 2014; Frizzo-Barker *et al.*, 2016; Mazzei & Noble, 2017; Shin, 2016; Wamba *et al.*, 2015). This study is directed at further characterising the influence of components of business capability within the base research model described by Akter *et al.* (2016) and Wamba *et al.* (2017) (Figure 1 below). This is done such that dimensions that influence effective deployment of innovations or capabilities such as BDA within businesses can be characterised, in order to derive tangible recommendations for business to execute. This is done with the intent of facilitating BDA’s effectiveness in an organisation, which is expected to improve firm performance.

The boundaries of this study are delineated through inclusion of three research constructs which are proposed to be assessed, in addition to the BDA model described by Akter *et al.* (2016) and Wamba *et al.* (2017). The proposed organisational constructs focus on assessing the influence of three distinct capabilities within business on the effectiveness of BDA capability, as measured by market share and/or financial return. The focus of this study is thus an assessment of factors of business capability which influence application of BDA. The BDA model itself comprises of 1st, 2nd and 3rd order constructs, wherein 1st order constructs are granular level attributes of the BDA environment. These roll up to 1 of 3 core capabilities that define a BDA capability as 2nd order constructs, namely BDA Management Capabilities (BDAMC), BDA Infrastructure Flexibility (BDAIF) and BDA Personnel Expertise (BDAPE). These in turn roll up to the 3rd order construct of Big Data Analytics Capability (BDAC), which represents the entire capability at the organisational level. As illustrated in Figure 1, BDAC is directly related to Firm Performance (FPer) as an organisational output of the model. It is at this point

that the organisational constructs assessed are positioned, including Organisational Culture to BDA (OCBDA), Firm Dynamic Capability (FDC) and Data Driven Decision Making (DDDM).

The overall aim of this research is to gain a holistic understanding of whether FDC, DDDM and OCBDA, influence the success of deployment of BDA capability. This includes achieving a better understanding the influence of the individual, cumulative and moderating relationships, if any, between the constructs considered and the relationship between BDAC and FPer. Furthermore, this study aims to further characterise actions that should be taken by organisations to enable Big Data strategies, as inferred by assessing the impact of the constructs described in this study.

3.2. Research Questions

Research questions addressed in this study were considered as four distinct assessments, each of which was addressed by individual hypotheses. These are described below as individual research questions.

3.2.1. Research Question 1

Research Question 1: Is there a positive relationship between BDAC and FPer?

This research question was include to provide a confirmatory measure of the base assumption that BDAC correlates positively with FPer, as has been found previously through being part of the base model applied (Akter *et al.*, 2016; Kim, Shin, & Kwon, 2012; Wamba *et al.*, 2017; N. Wang, Liang, & Zhong, 2012). This was assessed through H_1 , which provides a baseline for subsequent analyses.

H_1 : Big Data Analytics Capability (BDAC) has a significant positive impact on Firm Performance (FPer)

3.2.2. Research Question 2

Research Question 2: Is there a direct positive relationship between FDC, DDDM and OCBDA respectively (independent variables) and FPer (dependant variable)?

Prior studies, including the authors of the base BDA model applied herein, have described links between organisational capabilities (strategy alignment and operational dynamic capability) and FPer (Akter *et al.*, 2016; Wamba *et al.*, 2017). This study applied a similar approach, but considered three different aspects of the organisational that spanned acceptance of and assimilation of an innovation (FDC and OCBDA) to assess imbedding of the BDA capability

(Elbashir *et al.*, 2011; Hill, 2003; Kuratko, Montagno, & Hornsby, 1990; Ravichandran & Lertwongsatien, 2005; Sher & Lee, 2004; Wang *et al.*, 2012), as well as application of the outputs of the capability in business (DDDM) (Baines, 2014; CEBMa, 2013; Davenport & Patil, 2011; Lavalle *et al.*, 2011; McAfee & Brynjolfsson, 2012; Popovič *et al.*, 2012). These were each considered with respect to FPer directly in order to ascertain the degree to which the organisation impacts upon the BDA model output without BDA. This was assessed through **H₂**, **H₃** and **H₄**. This describes the impact of the environment in which BDAC is situated on FPer independent of whether BDAC is present.

H₂: Firm Dynamic Capability (FDC) has a significant positive impact on FPer

H₃: Data Driven Decision Making (DDDM) has a significant positive impact on FPer

H₄: OCBDA has a significant positive impact on FPer

3.2.3. Research Question 3

Is there a direct positive relationship between FDC, DDDM and OCBDA (independent variables) on BDAC (dependant variable)?

The causal links between organisational influence and BDA capabilities are well documented in literature (Aker *et al.*, 2016; Alharthi *et al.*, 2017; Frizzo-Barker *et al.*, 2016; Wamba *et al.*, 2015). This study required that this be empirically asserted, thus the direct impact of FDC, DDDM and OCBDA as organisational characteristics on BDAC were assessed through **H₅**, **H₆** and **H₇**. This describes the impact of the environment in which BDAC is situated on FPer independent of whether BDAC is present.

H₅: FDC has a significant influence on BDAC

H₆: DDDM has a significant influence on BDAC

H₇: OCBDA has a significant influence on BDAC

3.2.4. Research Question 4

What are the combined impacts of FDC, DDDM and OCBDA (independent variables) on the BDAC – FPer (dependant variable) relationship, and do these independent variables moderate that relationship?

Given that FDC, DDDM and OCBDA are capabilities within an organisation which are, by their nature and mode of interaction, they are anticipated to impact on FPer through their direct

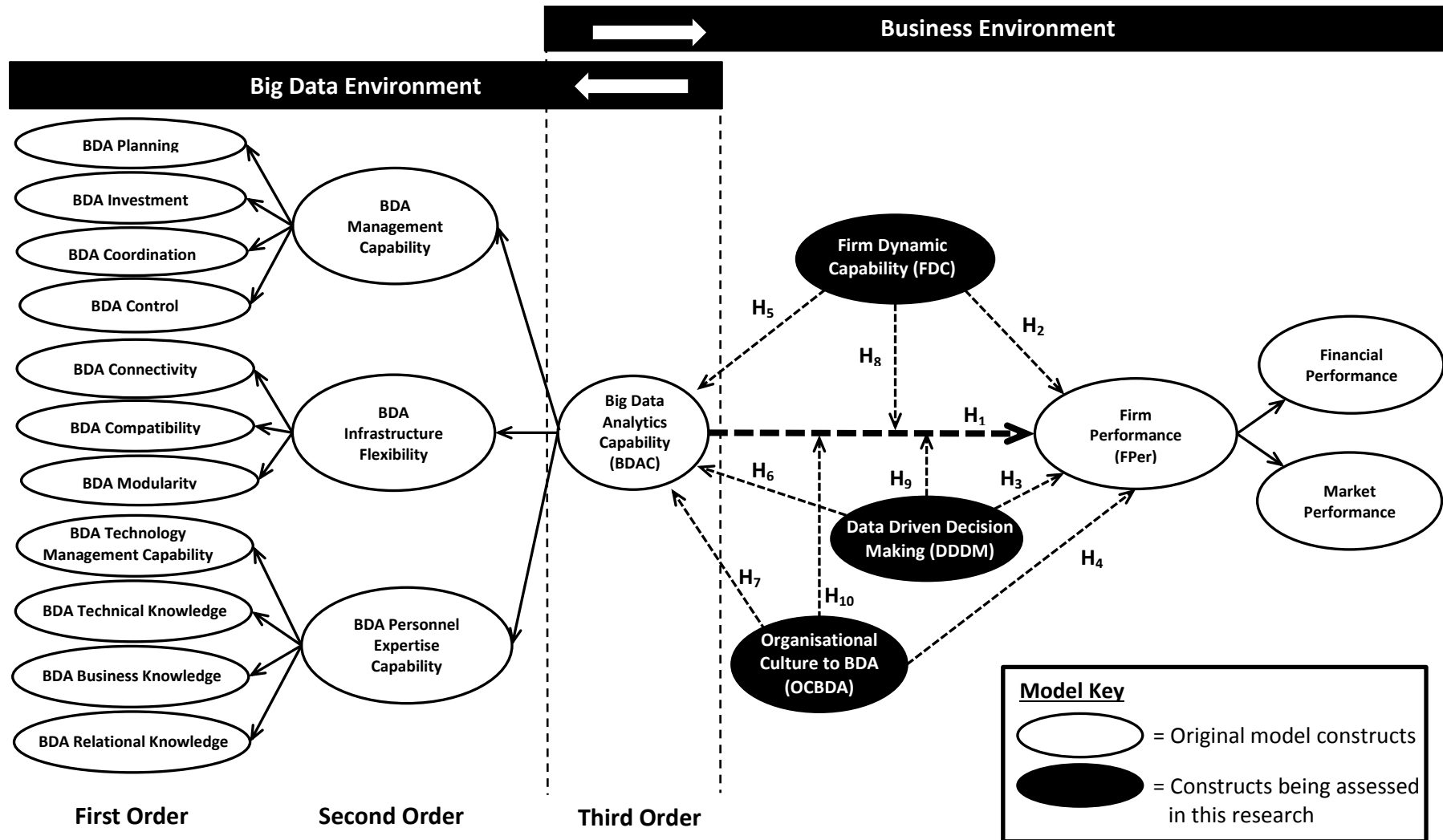
impact enabling BDAC to function effectively. Investigation of this thus assesses the direct relationship between each organisational construct with BDAC, and considers incorporation of it in the organisation, as well as application of it by the organisation. In so doing it also assessed for the impact on FPer through assessing potential for a moderating effect by these organisational capabilities. This thus assesses whether BDA as a capability and as a function which can deliver FPer is impacted by the organisational environment, and thereby assesses whether consideration of this aspect is valid when considering application of BDA by business leaders. This study assessed this relationship by investigating the combined impact of FDC, DDDM, OCBDA and BDAC as organisational capabilities (independent variables) and FPer (dependant variable) through **H₈**, **H₉** and **H₁₀**.

H₈: FDC has a significant positive moderation effect on BDAC enabling FPer

H₉: DDDM has a significant positive moderation effect on BDAC enabling FPer

H₁₀: OCBDA has a significant positive moderation effect on BDAC enabling FPer

Figure 1: Base research model applied in the study



Chapter 4: Research Methodology

“The real purpose of scientific method is to make sure Nature hasn’t mislead you into thinking you know something you don’t actually know” – Robert M. Pirsig

4.1. Introduction

As was described in Chapter 2, there is a developing body of literature which describes the influence of both organisational capabilities, and internal environment of an organisation on the ability to assimilate and the effectively capitalise on a capability such as BDA. McAbee *et al.* (2016) have detailed that, though fundamental to advancing the field, an over emphasis on deductive and theoretical literature persists in this arena. There is thus a need to both focus on the influence of the organisation on the implementation of BDA capabilities in business, and to do so by substantiating the causal relationships that have been asserted between application of Big Data environments and organisational capabilities. In this vein, this study has sought to characterise to which a set of constructs that describe aspect of organisational capability would impact on the relationship between BDA capability and FPer. Furthermore, the study endeavoured to assess cross relational attributes between the constructs to determine what their relative impact on the BDA – FPer relationship was.

4.2. Research Methodology and Design

As described by Saunders and Lewis (2012), the assumptions of a research philosophy inform a research strategy and resulting data collection methodology to be applied. These characteristics are describes hereafter starting with the broad concept of research philosophy, and narrowing down from there toward the methodologies applied in the study.

4.2.1. Research Philosophy and Approach

The intent of the study was to move toward empirically substantiating factors that influence the causal relationship that has been asserted between BDA and FPer. Furthermore, it was to assess whether this relationship was moderated by select organisational traits and capabilities. Given this, the approach taken the research philosophy governing the study was centred in positivism (Saunders & Lewis, 2012; Straub, Boudreau, & Gefen, 2004; Wamba *et al.*, 2017).

This assumes that phenomena are objective, can be expressed or measured through assessing causal relationships in data, and thus allows for findings to be asserted rather than postulated (Straub *et al.*, 2004). This is beneficial as it allows for empirical assessment and validation of assumed positions, which is necessary for this environment. It does however also result in nuances in interaction being lost, as in order for such assessments to be made, parameters are imposed through the research methodology which may limit the identification of factors not directly sought.

This approach was substantiated by the requirement for parameters of the BDA – FPer relationship to be practically attainable and measurable. Given the complexity and number of factors influencing the BDA environment, an ‘entanglement’ view of BDA was assumed, wherein several interlinked dimensions were assessed (Kim *et al.*, 2012; Wamba *et al.*, 2017). Despite this, not all possible parameters could be addressed. An approach of gathering data through structured surveys as a quantitative method was selected to facilitate measurement of the environment whilst addressing as many aspects as were practically feasible (Akter *et al.*, 2016; Saunders & Lewis, 2012; Straub *et al.*, 2004; Wamba *et al.*, 2017).

This study intended to expand the understanding of the influence of organisational capabilities on the ability of BDA to deliver value. This applied the base position that implementation of BDA in an organisation should be indifferent to implementing any new technical capability in a business, and thus did not require that it be conducted as an exploratory study. Rather, this informed the need for it to be a descriptive study that should apply existing constructs and measures to further characterise the BDA as a new capability within business (Saunders & Lewis, 2012). As with all research of this type, numerous methodologies may be applied to collecting and analysing data, each with their own assumptions (Bryman & Bell, 2007; Saunders & Lewis, 2012; Zikmund, Babin, Carr, & Griffin, 2012). Given the existing theoretical models available in literature, and the relatively direct comparisons being drawn between the proposed artefacts, the study may be further described as being deductive; as it applies existing and well documented theoretical positions to further assess the characteristics described previously (Saunders & Lewis, 2012).

4.2.2. Research Type and Strategy

As the components assessed were anticipated to be both individually and collectively influential, and as BDA management capability (BDAMC), BDA infrastructure flexibility (BDAIF) and BDA personnel expertise (BDAPE) have previously been shown to be key components of a

firms BDA capability and FPer (Akter *et al.*, 2016; Wamba *et al.*, 2017), a quantitative approach was applied to measure the interactions. This required that quantitative data be collected to ensure objectivity and to empirically characterise the relationships being assessed (Harwell, 2011). In so doing, the study applied a deductive approach. This assumes that the constructs and theories being applied are descriptive of the parameters within the environment, however not all parameters can actively or practically be assessed through this method, which means results must be bracketed within context of their scope. This does however allow for theoretical positions that are assessed to be asserted and validated to confirm their impact.

Research questions, which were informed by literature, aimed to interrogate the validity and extent of theoretical assertions and postulated relationships (Agresti & Franklin, 2007). The investigation of research questions through quantitatively attained primary data assumed the positivist paradigm that the reality of the BDA environment of an organisation can be measured (Orlikowski & Baroudi, 1990). Primary data in support of this required completion of online surveys by appropriately skilled persons in the field of BDA. These attributes describe the base of methodology applied in this study (Zikmund *et al.*, 2012).

4.3. Population

The population relevant to study can be described as all organisations which have an analytics infrastructure which houses at a minimum 1TB of data within a consolidated or directly accessible database to qualify as a Big Data system (Lee, 2017). Although a relatively limited number of organisations in South Africa were anticipated to qualify as having Big Data platforms, the statistics for the entire sampling frame with all potential members of the potential total population is not available as this is not a well-studied area. Furthermore, given the non-conformity of many technical characteristics within the Big Data environment, this is not a measure which one would easily or accurately be able to be measured given the timeframe.

A range of maturity of organisational capabilities, degrees of BDA implementation, as well as perspectives of several levels of experience and vocation within organisations was actively sought to provide a balanced view of the environment. These included area of expertise, division of the organisation (analytical, marketing, sales etc.), exposure to BDA as a function and seniority within organisations. Strata of seniority were characterised by the respondents' relative position within an organisation, such as CxO, junior or senior management. Furthermore, strata were sought according to business unit operational capability to separate

technical, analytical and business roles. This approach actively broadened the potential population sought by the researcher. The constructs being assessed were also expected to be present within all organisations, thus these were also deemed to not be limiting factors.

Though a broad range of personnel within the organisation were sought, their attributes were relevant to population, as a relatively high level knowledge and insight of business capabilities and deployments was required to appropriately respond to the research survey. Given this, characteristics relevant to limiting population included seniority, position, experience and relationship to BDA itself. Despite these conditions, a sufficiently large sample was expected to be achievable for this study.

4.4. Unit of Analysis

The unit of analysis of the proposed study was the companies or organisations which have a sufficiently large data set to allow for a BDA environment. In correspondence with this, questions posed related to the traits of the organisation, even though they may be reported on from the perspective of an individual. Furthermore, as the study aimed to elicit generalizable traits, this study assessed traits of organisations in general, and not a single specific organisation. In order to limit proliferation of erroneous assertions from data, consistency between the unit of theory, unit of measurement and unit of analysis were ensured (Mason, 2002; Zikmund *et al.*, 2012).

4.5. Measurement by Survey Questionnaire

As was described by Saunders and Lewis (2012), the assumptions of an adopted research philosophy inform the research strategy and data collection methodology applied. The primary determinants in this research were the research questions and study objectives. A single phase survey method was applied in collection of primary data in relation to the research questions posed (Chapter 3), the intended scope and timelines available. The survey was applied as a predetermined, self-administered online questionnaire. Though this carries significant benefits in that it allows for a large and dispersed sample to be reached, it also includes concerns and limitations described further through Section 9.1.4. As the survey considered accumulative insight from respondents which referred to organisational capabilities in implementing BDA, and its resulting impact on FPer at a point in time, the study was cross-sectional in nature (Saunders & Lewis, 2012; Zikmund *et al.*, 2012). This thus applied a mono-method wherein a single measure is used in analyses. Though practical as it enables direct assessment of the issues

being addressed; it does introduce bias as only a single measure is used (Saunders & Lewis, 2012). For the purposes of this study, application of mono-method was appropriate given that it had previously been validated, and that it afforded the direct relationship in question to be addressed, and for theoretical assertion on it to be empirically investigated. Furthermore, given that the time period available, as well as access to the research sample for the study was constrained, a mono-method was deemed appropriate.

4.5.1. Survey questions

Questions related directly to variables applied were extracted from literature in order to apply previously characterised and vetted questions and scales, which would limit complexity given time pressures and scope of the study. These were sourced in relation to the specific constructs being assessed, which themselves were directed at resolving the research questions described in Chapter 3. Questions included in the survey were intended to collect primary data which would afford objective confirmation of conditions pertaining to the environment and capabilities within organisations and how this influences BDA, whilst also endeavouring to limit bias and variability in responses and enabling reproducibility of results (Bryman & Bell, 2007; Harwell, 2011; Zikmund *et al.*, 2012).

The core research model (white circles of Figure 1) previously described by Akter *et al.* (2016) and Wamba *et al.* (2017) was applied as a basis in the study. That model provided a majority of the questions applied, as presented in Appendix 1, Table 21. In addition to the base questions applied by Akter *et al.* (2016) and Wamba *et al.* (2017) in assessment of FPer, firm market performance included further questions described by Wang, Liang, and Zhong (2012). Further constructs assessed in the study (black circles in Figure 1), were independently derived from literature. These included questions pertaining to OCBDA (Hill, 2003; Kuratko *et al.*, 1990), DDDM (CEBMA, 2013), Firm Dynamic Capability (FDC) and Firm Adsorptive Capacity (FAC) (Elbashir *et al.*, 2011; Ravichandran & Lertwongsatien, 2005; Sher & Lee, 2004; Wamba *et al.*, 2017; N. Wang *et al.*, 2012).

The research questions applied were initially compiled to a list of 120 questions. These questions were then scrutinised to avoid duplication, as well as to ensure all required aspects were covered. Through this process, questions were gradually removed which allowed for shortening of the total survey length. This was intended to both improve the likelihood of survey completion, as well as to improve the likelihood that answers would be provided with due consideration by respondents. Survey content and length was deemed appropriate at 72

questions (Appendix 1, Table 21). An introductory section was included which posed screening and classificatory questions to inform descriptive groupings of responses and to ensure appropriate capability of respondents. This also allowed for multiple levels of respondents to be characterised from data.

In each case appropriate scales were attached to questions to ensure appropriateness of potentially answers, described further in the following section. These completed questions and answers and scales were then transcribed into SurveyMonkey™ as an online survey tool.

4.5.2. Measurement Instrument

Given that the basis of the research model and associated research questions were those previously described by Akter *et al.* (2016) and Wamba *et al.* (2017), the parameters applied therein were duplicated in this study. Bar the introductory section which described respondents and their organisations, all questions were measured using a 7-point Likert scale, including both the questions applied by them in the base model, as well as the added constructs, allowing uniformity in data through the study. This is an appropriate methodology given the quantitative approach and the need to design the questionnaire based on validated instruments (Agresti & Franklin, 2007; Saunders & Lewis, 2012). This approach allowed for respondents to grade their perceptions of each of the questions by responding to the question according to a seven point scale of strongly positive to strongly negative, with three variations on that theme depending on the appropriateness of answers relative to the English language question that was asked.

It should be noted that although the wording in the scale varied given terminology in each specific question, the order of the scale from positive to negative remained consistent through the questionnaire in order to avoid inaccuracies of response error. Furthermore, questions and answers applied were extracted from previous research analyses, thus were assumed to have been well vetted and curated for application in research studies. Credibility of research data was strived for through maintaining objectivity in question set construction, and by being cautious to avoid errors and bias (Bryman & Bell, 2007). It should also be noted that the introductory section did not conform to the 7-point Likert scale as previously mentioned, as this was not appropriate to the questions posed. Rather individual scales pertaining to the question type were applied in this section, each of which is presented in Appendix 1, Table 21.

Pre-testing and assessment of the questionnaire prior to distribution in market is considered in further detail through Section 4.7.

4.6. Data Collection

Data was gathered via a single sample, self-administered cross-sectional assessment of organisations over a two month period (01 August to 30 September 2017).

4.6.1. Data capture by online survey

As in all research endeavours, the tools used in data collection impact on the questions which can be asked, as well as the data which one can collect. Though online surveys were initially viewed as a potential barrier to research by Kiesler and Sproull (1986), the targeted population were expected to be highly proficient in online and computer aided systems. Given this, the application of online survey was expected to pose no negative impact on respondent's ability to complete questionnaires. Rather, application of online survey was expected to allow for increased ease of distribution and completion of surveys by respondents who were expected to be highly dispersed through market, and thus hard to reach.

Online surveys were selected as the primary means of data collection as they allowed for efficient collection of data from a broad, dispersed and difficult to reach sample population. This method also supported the potential for a comparatively larger sample to be collected in a shorter period of time than is possible for in person assessments. This was a necessity given the anticipated challenge of accessing sufficient appropriately skilled respondents, given the relatively recent proliferation of BDA in the local market, and the expectation of it being a highly dispersed skill set.

In all cases the survey was distributed as a web link to the online survey, hosted and administrated within SurveyMonkey™. This allowed for ease of access by respondents as well as ease of data collection and survey setup and curation by the researcher. Furthermore, this afforded the landing page of the questionnaire to be a cover note that indicated an outline of the study and conveyed key notices such as anonymity of respondents, and an expected time requirement of approximately 15 minutes (on average it took 13 minutes for respondents to complete the survey).

4.6.2. Sampling method

Given the limited potential pool of appropriate candidates, as well as it being challenging to reach them due to them being dispersed through multiple companies, as well as them not easily reached through a central or professional body, as one does not exist, the sampling approach was two parts.

The initial part was directed primarily by non-probability purposive sampling, which was used to select for potential respondents to whom the survey could be sent. Therein, any persons in appropriate or associated positions were sought by the researcher directly, but also included an aspect of convenience sampling (Etikan, Musa, & Alkassim, 2016). This was done firstly, through personal and professional networks of the researcher and the research supervisor, both of whom being active in the field and having built up contacts over time. Selection was based on experience, expertise and position within an organisation which fitted criteria of having a BDA capability, and thus which remained consistent with measurement of the unit of analysis. As respondents may have had the survey forwarded to them without direct contact with the researcher, these parameters were further verified in the introductory portion of the survey through screening questions.

Application of this method provides the potential to introduce an aspect of sampling bias as a result of the sample potentially not being a true reflection of the entire population (Etikan *et al.*, 2016; Saunders & Lewis, 2012; Zikmund *et al.*, 2012). This was avoided as much as possible firstly by canvassing as broadly as possible across networks, secondly by ensuring multiple industries and business types were represented, thirdly by assessing various professional backgrounds and finally but expanding beyond direct networks, thereby ensuring sampling was not limited to convenience samples. Furthermore, given the limited total sample expected, as well as the reach imposed by intended distribution channels, sufficiently broad distribution was assumed for appropriate sampling of the potential population.

The second dimension of sampling included an aspect of snowball sampling, as persons contacted in the initial purposive stage were asked to pass the survey on to further people in their own networks and/or businesses who would be able to qualify as respondents. Given the dispersed and often hard to reach nature of the sample pool, this was done in order to distribute questionnaires and reach as many appropriately qualified persons as was possible (Saunders & Lewis, 2012). Though initial purposive screening by the researcher was not possible in this scenario, in-survey screening questions were included to identify respondent appropriateness. Though the researcher was unable to discern which responses were of purposive or snowball sampling respectively as all were anonymous, it was anticipated that though a bulk would have come from direct communication with respondents, a relatively large number would also have been as a result of snowball sampling.

Although application of this technique facilitated reaching the dispersed population, it's application does introduce a further potential of sampling bias, given that initial respondents tend to nominate colleagues or persons in their networks who may share similar traits and characteristics, thereby limiting representativeness (Zikmund *et al.*, 2012). Given that this was only a portion of the sample, and given the spectrum of respondents the relative impact of this, the concern was deemed to be minimal.

4.6.3. Survey distribution

Collection of a sufficiently large amount of data is central to drawing objective, generalizable and reproducible results. Given the requirement to maintain statistical validity, one runs the risk of encountering samples sizes that themselves are not significant, which would cast aspersions of results (Agresti & Franklin, 2007; Saunders & Lewis, 2012). This was mitigated by making the survey available on multiple platforms, and by making the format consistent and easily interpretable (common structure of easily selectable answers from common and intuitive ranges). Further consideration of the challenges of application of online surveys for data collection, and mitigation thereof is presented in Section 9.1.4 of Appendix 1.

As described in Section 4.6.1, distribution was carried as a dispersal of a web link to an online survey. As described in Section 4.6.2, potential respondents were targeted through direct contact, by referral from persons contacted, as well as by dispersal through online platforms in public domains. This included primarily direct communication with potential respondents within the researchers personal and business network (researcher works in the BDA community in South Africa) who were determined to have appropriate skills to answer the survey themselves. These respondents all received personalised communication from the researcher to support survey completion. Appropriately skilled persons were also actively sought by researching positions on LinkedIn, and actively making contact with the individuals found thereon. This was also done at conferences in the field of BDA held in Johannesburg South Africa, detailed in Appendix 1, Table 22. Experience, position and company with which the individual was associated were applied as criteria in these searches.

Furthermore, people within the researcher's network who were in appropriate positions, or appropriate organisations, that allowed for appropriately skilled respondents to be reached were asked to distribute the survey on the researcher's behalf. This was done according to a description of appropriate respondent by the researcher. Internal screening questions within

the survey were included to verify respondent appropriateness it not personally vetted by the researcher before time of distribution.

General distribution on public platforms primarily involved posting social media platforms including LinkedIn, WhatsApp, WeChat and Facebook. In order to drive notification and uptake of social media posts, they were actively 'liked', re-posted, commented on and requested to be liked and commented on by associates. This was done as those platforms display items with more interest to more people more frequently as a function of their operation, thus these actions were intended to promote the potential for viewership and engagement through those platforms. In order to describe specific examples, surveys were posted in dedicated BDA groups non LinkedIn after gaining authorisation from group admins on several separate group pages. These are summarised in Table 22 of Appendix 1, and show potential for distribution to hundreds of thousands of possible respondents. This also included personal recommendation for the surveys in reposts by the programme coordinator of the MIT Big Data and Social Analytics group, as well as the administrator of the Big Data and Analytics pages.

Distribution was also furthered by the researcher's supervisor, who reached out to his own network, through both sharing of the survey on social media pages on several occasions, as well as several public lectures and workshops associated to BDA which were given by the supervisor, and where attendees were associated to the field. Furthermore, this was also done through engaging the remainder of the MBA cohort, as well as additional colleagues in the Global Trilateral MBA programme (GTMBA) to access their networks. These avenues yielded limited results as only a handful of these colleagues has access to appropriately skilled persons.

4.7.Survey Questionnaire Pre-testing

In order to validate assumed correctness or ease of understanding of questions in the survey questionnaire, both in their own respect and as components of constructs, pretesting of questionnaires was performed.

4.7.1. Questionnaire screening

Initial screening was performed by the researcher and the researcher's supervisor. This check was primarily a validation of design and flow of questions and confirmation that they both addressed the research questions posed. The GIBS research ethics Committee then assessed whether the survey violated any ethical standards; clearance for the research questions

prepared is attached in Appendix 2. The primary action post this stage was to reduce survey length from 97 to 72 questions, as well as including country in which respondents work, the final result appended in Appendix 1, Table 21. Country was included as there was potential for international exposure which may carry differences in perceptions relating to organisational capability and how this may relate to inception of capabilities such as BDA. This dynamic is described by cultural dimensions of Hofstede (n.d.). Once these issues had been resolved, the questionnaire was transcribed into the online survey tool, SurveyMonkey™ that afforded further minor formatting, such as adding progress bars, drop down menu's and visual aesthetics.

4.7.2. Sample respondent pre-test

A pilot survey was then run by distributing the online version of the questionnaire to 14 respondents who assessed inconsistencies in flow, legibility, grammatical errors, ease of understanding relative to researcher's intent by a person who did not have prior background about the research project.

Feedback from pilot respondents was assessed to ensure clarity and consistency of questions asked, as well as to validate the questions and to avoid any biases (Saunders & Lewis, 2012). This revealed no major concerns about the questionnaire design, other than to make it explicit on each page that if answers to questions were not known by respondents, those could be skipped. This was done both to avoid erroneous responses and to promote completion of the survey, even if some answers were not known, as some respondents felt that they were not sufficiently skilled in all aspects of the survey (which covered a spectrum from business to technical questions).. One question was slightly reworded to reduce potential of confusion as it was negatively worded relative to surrounding questions which may have been confusing to some respondents. Furthermore, the initial time estimate to complete the survey was revised down from 30 to 15 minutes given respondent's results.

4.8. Survey Data Collation

After the minor edits described above were complete, the survey was actively distributed as described in Section 4.6.3 over a 2 month period.

Though the researcher was unable to determine the entire number polled given the exposure from multiple platforms, it can be relatively safely assumed that the response rate was well below the average rate of for online surveys of between 15% - 30% (Fricker, 2008), given that

even if only 1% of only the 351,094 members in the Big Data and Analytics group on LinkedIn viewed the survey, that would be 3,510 people polled, with a resulting response rate of 3% given 109 responses received. This illustrates a much lower than anticipated response rate received in the study. This may be a measure of some of the concerns of online surveys reported by Evans and Mathur (2005). Alternatively this may be a measure of apathy among the sample cohort. Only 51 of these 109 respondents were complete in totality thus further curation of data was required to determine applicable sample.

4.8.1. Background on managing missing survey data

Though more acute in longitudinal studies than cross-sectional studies, missing data or incomplete questionnaires is a common ailment of research studies which has been addressed in many forms, ranging from deletion of data to substitution of answers (Graham, 2009; Honaker & King, 2010; King, Honaker, Joseph, & Scheve, 2001; Schafer & Graham, 2002; Scheffer, 2002). The need for complete data is due to typical analytical techniques requiring completeness of data (rectangular or non-patchy data) for statistical correctness (Graham, 2009; King *et al.*, 2001; Scheffer, 2002). Though the problem of how to deal with missing data, once it has been deemed impossible to recover those missing values, has been actively addressed as early as Rubin (1976), it remains a topic under development (King *et al.*, 2001; Scheffer, 2002).

Traditional approaches to managing missing data include case deletion and mean imputation. King *et al.* (2001) reported that approximately 94% of reports reviewed used listwise deletion to eliminate entire observations when a variable remained missing after filling in guesses, which resulted in loss of about a third of data. Given the already limited sample of this research, that is not a desirable result, thus imputation has been opted for to 'complete' surveys. More recently, interest on Regression Imputation, Imputation using the Expectation – Maximisation (EM) algorithm (both of which use Single Imputation) and more recently, Multiple Imputation has become available in some statistical packages (King *et al.*, 2001; Scheffer, 2002).

As is described by King *et al.* (2001), Schafer and Graham (2002) and Scheffer (2002) respectively, there are three 'Missingness Mechanisms'. Missing Completely at Random (MCAR) refers to data that was collected randomly, and doesn't depend on the variable of interest, or any other variable in the dataset, so is both missing at random, and observed at random. Not Missing at Random (NMAR), or informatively missing is when the Missingness

mechanism depends on the actual value of the missing data and is the most difficult condition to model for. Finally, Missing at Random (MAR) refers to data that is missing as a condition of some 'X variable' in the data set, but not the 'Y-variable' of interest. More specifically, it means that a respondent's probability to respond may be related only to their own set of observed items, which may change from one respondent to another (Schafer & Graham, 2002). Given this, data is not really missing at random at all, but rather not included.

Given that the missing data in this studies data is expected to be due to the breadth of subject matter covered, it is likely both to be a factor of respondents not having sufficient exposure or breadth of knowledge of all constructs (X variables) and thus the messiness of this study is a factor of MAR. There is no way to test whether MAR holds true in a given data set other than through obtaining follow-up data from non-respondents (Schafer & Graham, 2002). This was not possible in this study as all respondents were anonymous, thus the assumption of MAR is presumed to hold true.

4.8.2. Management of missing survey data in this study

As described by Scheffer (2002), though multiple imputation methods work well up to 25% of data missing, single imputation methods do work well for MAR, but only when there is less than 10% of data missing, if interested in the mean. This drops to 5% of missing data when variance structures in data are important, then don't use these methods if the data is more than 5% missing. Application of an average to fill in missing data is a widespread practice, though its properties remain largely unstudied and its practice doesn't even have a well-recognised name (Schafer & Graham, 2002). Though its application has theoretical issues such as the potential for introducing bias, increasing variance in a scale, and reducing reliability, preliminary investigations suggest that the method provides reasonable results (Schafer & Graham, 2002). This was largely validated by Scheffer (2002) who considered the impact of up to 50% of data being missing on mean and standard deviation. Due to this, 50% missing data per questionnaire is set as the maximum and averaging applied as a base mechanism. They did however state that with missing data as high as 50%, multiple imputation should be applied as the method as this is far more accurate.

Missing data were scattered through the data set and questions were not uniformly missing (i.e. whole questions unanswered by all respondents). Due to this, the issue of mean substitution not providing sufficiently representative variance on the Y-axis reported by Schafer and Graham (2002) was considered to be of limited impact. This was further reduced

by substituting mean values for missing data where respondents were grouped into multiple sub-groups based on descriptive traits that were relevant to the construct in which the question for which missing data was in (e.g. industry of respondent for general responses, experience of respondent for the extent of implementation of a BDA capability or whether the respondent was in an analytical, business user or management role to account for technical questions concerning analytics and systems capabilities) (Hair, Black, Babin, & Anderson, 2010; Schafer & Graham, 2002; Scheffer, 2002).

4.8.3. Data coding for imputation

To allow for issues of missing data to be resolved, English language answers needed to be converted to numeric values such that missing data could be imputed mathematically. The data collected through the survey was encoded by converting English language answers (i.e. strongly agree through to strongly disagree in the 7-point Likert scale options provided) to numeric values. The table illustrating the combination of answers to numbers is provided in Appendix 3 (Table 24).

4.8.4. Imputation of missing survey data by averaging

As informed by King *et al.* (2001), Schafer and Graham (2002) and Scheffer (2002) and described further in Appendix 3, missing data in surveys with at least 50% of questions completed were filled in through imputation using averaged responses. More specifically, this was done by averaging responses per question from a group of respondents which portrayed similar characteristics to the respondent. Importantly, the relevant group was determined through filtering for descriptive traits (Section 1 of the questionnaire in Appendix 1, Table 21) that best related to the question being asked. For example, if the question related to competitors, groups were filtered by industry (such as telecommunication or insurance respectively) whereas if it asked for technical details, groups were filtered by type of role (managerial, executive, analyst, systems engineer or user of data) to provide to most appropriate grouping from which responses could be averaged.

4.8.5. Final Sample Size

51 respondents (47%) of the sample completed all responses, whilst the remainder (56) of the total 109 respondents left out one or more answers. 29 of the responses had more than 50% of questions completed, thus they had the unanswered questions completed as described above, bringing the total sample size to 80 responses, all of which answered affirmatively to

the screening question, and thus all could be applied. The final coded data set is presented in Appendix 3, Table 28.

Raw data files as well as results of imputation and all subsequent analyses were provided in electronic format attached to the hard copy of this document.

4.9. Analysis Approach

Data collected was quantitative, categorical and ordinal in nature. As data was collected via an online survey tool, it was already collected in digital format that could be exported directly to Excel and SPSS for further analyses. It was from this basis that the data was coded as described above. Analysis of data was conducted using both descriptive and inferential approaches. Descriptive assessment were primarily applied to describe and characterise the sample set, while inferential statistics was used to test the hypotheses developed from the research questions.

4.9.1. Descriptive statistics

4.9.1.1. Assessment of respondents and their organisations

Descriptive statistics were used to describe the central tendency, relative contribution and spread of the data set collected. Within the total data set, this entailed assessment of the number of total responses and an evaluation of the percentage completed, as well as their relative contribution within the data applied. This included the total data (including the responses that were disqualified) as well as responses which had missed or not-completed questions imputed, as described in Section 4.8.4. These assessments included making use of the mean and standard deviation respectively for the continuous variables (strongly negative to strongly positive). Assessment of categorical variables (those applied in Section 1 of the questionnaire, Table 21 in Appendix 3) included frequency and percentage frequency to describe the data meaningfully. Where applicable, graphical representation was used to illustrate data, including bar, line and pie charts.

4.9.1.2. Shapiro-Wilk test for normality

Parametric analyses require that data be normally distributed and are typically influenced by data which are outliers, or which deviate significantly from the mean of the data set (Bryman & Bell, 2007; Hair *et al.*, 2010). Though Kurtosis and skewness of data are representative of the distribution of data, this can be more effectively measured through the Shapiro-Wilk test for normality, and allows for the assumption of data distribution to be verified statistically

(Shapiro & Wilk, 1965). In the absence of normality, non-parametric analyses must be applied, this this parameter is highly influential in the correct determination of results. In this study, the Shapiro-Wilk test was applied ($p < 0.05$), wherein if the null-hypothesis was confirmed ($p < 0.05$), then data are not normally distributed (Razali & Wah, 2011; Shapiro & Wilk, 1965). This is reported for the factor constructs derived in this study in Table 16.

4.9.2. Internal reliability of data and constructs (Cronbach's alpha)

4.9.2.1. Introduction

The assessment of a quantity or measure of interest in relation to a population is often impossible to measure explicitly (Bland & Altman, 1997). In place of an explicit measure, researchers ask a series of questions and combine those answers into a single numerical value which is representative of the desired measure (Bland & Altman, 1997). When used to form a scale, the items assessed or questions asked, should all measure the same thing and thus they should be correlated to one another (Bland & Altman, 1997; Cronbach, 1951). The presence of such internal correlation provides reliability in the measure, whereby the measurement instrument is expected to produce consistent results if those it were to be used on other occasions (Cronbach, 1951; Saunders & Lewis, 2012).

Cronbach's alpha is most commonly used coefficient for assessing internal consistency, and is described by the following formula;

$$\alpha = \frac{k}{k-1} \left(1 - \frac{\sum s_i^2}{s_T^2} \right)$$

Where k represents the number of items, s_i^2 the variance of the i^{th} item and s_T^2 the variance of sum of all the items (Bland & Altman, 1997). It is typically used where multiple Likert questions are present within survey data that are used to form the construct or scale (Kline, 2016), as was the case in this study. It is also commonly used to account for consistency where data reduction techniques such as principle component analysis is applied (Kline, 2016).

4.9.2.2. Assessments conducted

Internal consistency of the constructs applied (Table 21, Appendix 1) was confirmed in two ways. The first was through Cronbach's alpha as the correlation coefficient, and the second was through Factor Analysis on each of the main sections or constructs of the survey. Section 1 was not included in the assessments as it contained questions that were descriptive of the sample set, not Likert intended to be applied in constructs. Though factor analysis is

considered in Section 4.9.4, the following describes application of Cronbach's alpha. Descriptive assessments of construct were included in Table 2 below, along with detailed per-question Cronbach's alpha and descriptive results presented in Appendix 4, Table 29. Cronbach's alpha assessment allowed for scaling if items were deleted. In each case, if $\alpha < 0.7$, questions were deleted one at a time with analyses redone per iteration, as described in Appendix 4, Section 9.4.2, and illustrated in Table 30.

Table 2: Summary Descriptive statistics of survey constructs

Descriptive Statistics									
Construct	N	Minimum	Maximum	Mean	Std. Deviation	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
OCBDA	80	2.63	7.00	5.1047	0.98995	-0.383	0.269	-0.311	0.532
DDDM	80	2.50	6.83	4.5688	0.92690	0.080	0.269	-0.231	0.532
FDC	80	1.83	6.67	4.3375	1.21320	-0.208	0.269	-0.883	0.532
BDAMC	80	1.88	6.41	4.3846	0.89773	-0.308	0.269	0.172	0.532
BDAIF	80	2.13	6.13	4.1672	0.94193	-0.087	0.269	-0.270	0.532
BDAPE	80	3.00	7.00	5.1792	0.90310	-0.430	0.269	-0.202	0.532
FPer	80	1.00	6.86	4.4554	1.14739	-0.543	0.269	0.733	0.532

4.9.3. Internal validity of data and constructs

4.9.3.1. Introduction

Validity of a measure refers to ensuring that the measurement instrument both accurately measures what was intended to be measured, and that it is reflective of what was intended to be measured (Adcock & Collier, 2001; Hair *et al.*, 2010; Johnson, 1997; Winter, 2000; Zikmund *et al.*, 2012). This is thus a critical analysis that provides confirmation that, along with reliability of measure, allows for the assumption that data collected is representative, and thus that it can be applied within analyses. As is described by Winter (2000), the measure of validity should not be confused with precision, which describes measurement to n-decimal points. This is because a measurement should be relative to the purpose it is being applied, and the more 'precise' the scale, the more difficult it is to achieve high 'validity' (Winter, 2000). Quantitative studies such as this one are by nature concerned with 'construct validity', where the construct describes the hypothesis, concept or question, and that determines what data should be gathered, and how to go about doing so (Cronbach & Meehl, 1955; Johnson, 1997). Researchers applying this methodology actively cause or affect an interplay between data and construct to 'validate' their investigation (Cronbach & Meehl, 1955). This describes the need

for internal validity, wherein a test, such as a bi-variate correlation, is typically applied to verify whether a researcher is justified in concluding that an observed relationship is causal, and thereby indicate whether it can put forward as an assertion of that study (Hair *et al.*, 2010; Johnson, 1997; Zikmund *et al.*, 2012).

4.9.3.2. Assessments conducted

Internal validity of construct was confirmed in two ways. The first was through bi-variant correlations between each sample question, and the construct mean, further details towards which are described in Section 4.10. The second was through confirmatory factor analysis, discussed further in Section 4.9.4 (Hair *et al.*, 2010; Zikmund *et al.*, 2012).

4.9.3.3. Internal validity of data sets

The constructs that were derived from Cronbach's alpha analyses were assessed for internal validity as a further confirmatory step prior on further application in analyses. This was done to ensure constructs that were extracted from literature, or compiled from multiple sources retained validity. This was initially performed by Pearson's correlation, with all questions displaying significance within their respective constructs, and thereby confirming construct validity. These results are summarised in Table 13, and presented in further detail in Appendix 4, Table 32. Internal Validity was further validated post Exploratory Factor Analysis (EFA) through Confirmatory Factor Analysis (CFA), the results of which are presented further in Section 9.4.3.

4.9.4. Factor Analysis (FA) of data

4.9.4.1. Description of methods

Factor analysis is a technique by which the number of variables in a construct set can be reduced through statistically assessing interdependence between factors (Zikmund *et al.*, 2012). The primary purpose of doing so being the determination of the underlying structure among variables assessed (Hair *et al.*, 2010). Doing this allows one to simplify analyses, and is necessary as the need for increased knowledge about the structure and interrelationships of variables increases as the number of variables being considered increases (Hair *et al.*, 2010). Therefore, by identifying the variables which account for the greatest variance of the original variables applied, variance in results may be accounted for by selecting for the lowest possible number of key variables, which in turn allows for more simplicity in classifying and assessing responses (Hair *et al.*, 2010; Zikmund *et al.*, 2012).

In these analyses, factors are identified by forming a variate using the measured variables of the data set (Zikmund *et al.*, 2012). This is a procedure which allows for highly correlated variables to be grouped and labelled, re-labelled or to form new composite measures (Hair *et al.*, 2010).

Factor analysis may be separated into two categories, namely Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA). EFA is performed when the number of factors present in a set of variables is unknown, whilst CFA is performed when there is a strong theoretical expectation of the factor structure (both the number of factors and which variables relate to each those factors) before performing the analysis (Hair *et al.*, 2010; Zikmund *et al.*, 2012). CFA is a form of structural equation modelling which specifically considers measurement models (Brown & Moore, 2012; B. Williams, Onsman, & Brown, 1996). More specifically, it deals with the relationships between observed measures (*indicators*, such as reported items, scores, and ratings) and latent variables (*factors*) (Brown & Moore, 2012; B. Williams *et al.*, 1996). The intent of latent variable measurement models such as factor analysis is the identification of the number and the nature of factors which account for the variation and covariation among a set of indicators. Accordingly, the factor itself is an un-observable variable which holds influence over observed variables, and accounts for the correlation between those variables because they share a common underlying feature or construct (Brown & Moore, 2012). CFA can be used to examine the latent structure of a test instrument and verify the number of underlying factors present, and thus allows for it to validate constructs through providing compelling evidence of the *convergent* and *discriminant validity* of a theoretical construct (Brown & Moore, 2012; Zikmund *et al.*, 2012).

Within EFA or principle component analysis (PCA), Bartlett's Test of Sphericity matches an observed correlation matrix to that of the identity matrix, allowing for redundancy between variables to be identified, and the number of factors to be reduced (Hair *et al.*, 2010). In so doing it assesses the significance of the study, and describes the validity and suitability of responses in relation to the problem being addressed (B. Williams *et al.*, 1996). Kaiser-Meyer-Olkin (KMO) measures of sampling adequacy, and through this illustrates whether variables can be factorised (Hair *et al.*, 2010).

4.9.4.2. Assumptions of EFA

As described by Hair *et al.* (2010) and Zikmund *et al.* (2012), there are four assumptions made when analysing data by factor analysis. These include firstly that the variables are continuous

or at least ordinal, which is the case in this study given the data Likert scales applied. The second assumption is that there should be a linear relationship between all variables. Confirmation of this assumption in the data applied is provided through select scatter plots in Appendix 4, Section 9.4.4, Figure 4. The third assumption is that the data contain no outliers. These are confirmed through component scores being no greater than 3 standard deviations from the mean within Factor analysis results. The fourth and final assumption was that there is a sufficiently large sample size. This is a limitation of the study, and is a factor which could not be raised further than a sample of 80 respondents.

4.9.4.3. Assessment by EFA

This study applied EFA as a means of assessing construct composition through development of factors per section of the survey questionnaire. EFA were conducted on each section firstly to identify which questions loaded together to form factors or the construct. Secondly, it identified which factors did not have sufficiently high correlation, and thus did not load and could be removed, allowing for factor reduction of constructs (Hair *et al.*, 2010; Zikmund *et al.*, 2012). The Kaiser-Meyer-Olkin's (KMO) measure of sampling adequacy and Bartlett's test of sphericity were used for EFA analyses. Varimax rotation was applied and all iterations were set to maximum of 25 for convergence. Extraction of number of constructs per data set was determined using principal component analysis (PCA) as the extraction method with eigenvalues set at greater than 1, and was supported visually by Scree plots. Each construct had 80 responses, however question number varied depending on construct, as presented in Table 12.

4.9.4.4. Assessment by CFA

Results of EFA allowed for questions to be grouped into factors which relate to one-another, and which could be applied in subsequent analyses. Each of the initial survey constructs, and its resulting factors as described from EFA were evaluated by CFA to validate their composition by building a separate model of each construct's composition in AMOS. A standardized estimate model was applied, wherein both the latent variable and indicator were in standardized metrics ($M = 0$, $SD = 1$) (Brown & Moore, 2012), allowing for the model to be interpreted as the correlation between the indicator and the factor.

4.10. Assessment of the relationship between constructs

4.10.1. Correlation between constructs

4.10.1.1. Description of method

Correlation coefficients are used in statistics to measure how strong a relationship is between two variables, or rather how the variable co-vary (Agresti & Franklin, 2007; Hair *et al.*, 2010; Zikmund *et al.*, 2012). There are several types of correlation however the Pearson product-moment correlation is most commonly applied in linear regression as it determines the strength and direction of a linear relationship between two continuous variables (Kline, 2016). The test generates a coefficient called the Pearson correlation coefficient, denoted as ρ when assessing entire populations, and as r when assessing correlations in samples, as is the case in this study (Kline, 2016; Zikmund *et al.*, 2012). The result of the Pearson correlation coefficient is a set of values which may range from -1 to +1, where -1 describes a perfect negative linear relationship, 0 indicates no relationship, and +1 a perfect positive linear relationship (Agresti & Franklin, 2007; Hair *et al.*, 2010; Kline, 2016; Zikmund *et al.*, 2012).

When concerned with a sample of a population, as is the case in all analyses herein, the sample Pearson's correlation coefficient (r) is applied, according to the formula;

$$r = \frac{n(\sum x_i y_i) - (\sum x_i)(\sum y_i)}{\sqrt{[n \sum x_i^2 - (\sum x_i)^2][n \sum y_i^2 - (\sum y_i)^2]}}$$

Where n is the sample size, x_i and y_i are the single samples indexed with i (Agresti & Franklin, 2007; Hair *et al.*, 2010; Kline, 2016; Zikmund *et al.*, 2012).

Application of this requires consideration of five assumptions regarding study design, measurements to be made, and how the sample data fits the Pearson's correlation model (Hair *et al.*, 2010; Kline, 2016; Zikmund *et al.*, 2012). These assumptions are described hereafter in accordance with Hair *et al.* (2010) Kline (2016) and Zikmund *et al.* (2012), with the first assumption being that variables should be measured on a continuous scale. This assumption is met in this study as data were collected using Likert scales. The second assumption is that the continuous variables be paired. The third assumption is that there is a linear relationship between the two continuous variables, and the fourth that there are no significant outliers, both of which were assessed by plotting scatterplots of data. The fifth and final assumption is that data are normally distributed, which was assessed using the Shapiro-Wilk test for normality. This is due to Pearson's correlation coefficient being sensitive to

skewed distributions and outliers (Kline, 2016; Zikmund *et al.*, 2012). Failing these assumptions would require application of Spearman's rank correlation as a non-parametric alternative, thus it was also run with Pearson's correlations for comparative assessment (Kline, 2016; Zikmund *et al.*, 2012).

4.10.1.2. Assessment by correlation

Given that Pearson's correlations is parametric analysis tool, and is influenced by outliers (Kline, 2016; Zikmund *et al.*, 2012), both Pearson's and Spearman's correlations were run on data for parametric and non-parametric assessments respectively. Correlation will allow for the understanding of the degree of co-variance of constructs, or the degree to which constructs relate to one another, and thus illustrated the relationship between variables. This is necessary as it allows for confirmation of hypotheses in the study. The assumptions for correlation as described section 4.10 were not all individually assessed prior to factor analysis as, by the nature of factor analysis, the composition of constructs was anticipated to change (Zikmund *et al.*, 2012). Construct and question level descriptive were however presented through Table 2 and scatter plots Figure 4. In this study, the correlation analyses were at 95% confidence interval.

4.10.2. Assessment of the relative influence of constructs on firm performance through hierarchical multiple regression (HMR)

4.10.2.1. Selection of method

Initially, Structured equation modelling (SEM) was intended for application in this study in order to assess the relative impact of the constructs on BDAC and FPer, as described by (MacCallum & Austin, 2000), and as applied by Akter *et al.* (2016) and Wamba *et al.* (2017) in closely aligned studies that applied the same base model. Unfortunately, sample size is of significant impact on the ability to carry out such a study, particularly the number of respondents required increases with the number of constructs assessed (MacCallum & Austin, 2000; Schreiber *et al.*, 2006). Given these parameters, a minimum sample of 240 respondents was initially sought for this study. As is described by Knofczynski and Mundfrom (2008) though the size of the sample (n) contributes strongly to the 'value' of a set of statistical results, contention remains regarding the appropriateness of a sample size for a given statistical method and application. Knofczynski and Mundfrom (2008) furthermore describe that the sample size required for prediction is almost certainly larger than for testing for statistical significance. As SEM is typically reported as requiring $n \geq 220$, though this too varied (Akter *et*

al., 2016; Hair *et al.*, 2010; MacCallum & Austin, 2000), the sample acquired in this sample was considered to be insufficient for this type of analysis.

Regression analyses are comparatively simpler assessments with lower sample number prerequisites. Knofczynski and Mundfrom (2008) provide a detailed description of appropriate sample size for multiple linear regressions based on the level of squared population multiple correlation coefficients. These studies reported Adjusted R^2 that ranged from 0.56 to 0.64, which translates to an appropriate sample for prediction with 6 variables of between 40 and 85 respondents (at 0.7 and 0.5 respectively). Given this, the sample of 80 respondents available herein was considered appropriate for the purposes of the study, and HLR could be applied.

4.10.2.2. Description of method

Multiple regressions are appropriate for analyses where single metric variables (FPer) are presumed to be related to two or more independent variables (OCBDA, DDDM, FDC) (Hair *et al.*, 2010). The objective of multiple regression analyses is the prediction of changes in dependent variables in response to shifts in independent variables, typically achieved through the statistical rule of least squares (Hair *et al.*, 2010). Hierarchical linear multiple regression (HMR) is a means of assessing whether variables of interest explain a statistically amount of variance in a dependant variable, such as FPer, after accounting for all other variables present (Hair *et al.*, 2010). As described by Davison, Kwak, Seo, and Choi (2002), Hierarchical regression differs to single-level regression analyses in three significant ways. Firstly, the slope and the intercept are allowed to vary as homogeneity or regression lines is a hypothesis that is tested. Second, hierarchical regression includes additional parameters which correspond to the variance of the intercepts and slopes. Third, when intercepts and slopes do vary, those variables are added into Model 2 equations so that they account for variation of intercept and slope. This method thus provides a framework for comparison of models by building several regression models, wherein variables may be added to a model in subsequent steps (Hair *et al.*, 2010). This is done to determine whether the newly added variable results in a significant in the R^2 of the model, which describes the proportion of variance explained by adding the component to the model. This thus provides a stepwise assessment of the impact of a construct relative to other variables.

4.10.2.3. Assumptions of method

Prior to conducting HMR, assumptions of the method were tested. Assumption of sample size was verified through Section 4.10.2.1. Assumption of singularity was met by independent variables not being a combination of other independent variables. Assessment of correlations (Table 36) revealed correlation between factors, including those which were not isolated from the same construct, with the exception of some relationships within 'Accessible', 'BDA tech applicable' and 'BDA knowledge transfer', thus the assumption of multicollinearity was deemed to have been met (Hair *et al.*, 2010). No extensive outliers were found as investigated by scatter plot, which also enable confirmation of linearity and homoscedasticity (Hair *et al.*, 2010). Normality of data were assessed through Shapiro-Wilk (Table 15 and Table 16), however skewness and non-normality of data were identified. These were not extensive, thus linear regression was considered appropriate, though results needed to be considered relative to this finding.

4.10.2.4. Assessment of HMR

A two stage HMR was conducted with FPer set as the dependent variable. Each of the variables barring the one being tested were added in stage one, whilst the variables that related to the construct being assessed were added in stage two in each case. This process was repeated with the factors of OCBDA, DDDM, FDC and BDA respectively being added in separate HMR in stage two, and in each case, $p > 0.05$.

Chapter 5: Results and Analysis

“If you can’t measure it, you can’t manage it” – Peter Drucker

5.1. Introduction

This Chapter is dedicated to illustrating the results of this study. It begins by applying descriptive statistics to characterise the sample set and describe the data from the survey in order to provide context. Following this, statistical analyses are presented, as described through Chapter 4, which interrogate the constructs and address the research questions described in Chapter 3.

The primary aim of this study was to gain a deeper understanding of whether firm capability within the three constructs (described in Chapter 3) influence the impact of BDA capability on FPer. This includes understanding the influence of the individual, cumulative and moderating relationships, if any, between Firm Dynamic Capability (FDC) which includes Firm Adsorptive Capability (FAC), Data Driven Decision Making (DDDM) and Organisational Culture to Big Data Analytics (OCBDA) and the performance of BDA capability as measured by Firm Performance (FPer). These are on the relationship between BDA capability and FPer, as stated in Chapter 3. The overarching requirement of these analyses is this the confirmation of the expected positive relationship between BDA and FPer, and then the determination of first whether, and then the extent of influence of the constructs on that relationship.

5.2. Descriptive Characterisation of the Sample of Data

5.2.1. Raw sample size

As is detailed by Knofczynski and Mundfrom (2008), a minimum acceptable sample size for multiple regression analysis with three independent variables is 127. For the purposes of this study, a sample minimum of 250 was initially targeted in order to allow for multivariate analysis, as performed by Akter *et al.* (2016) and Wamba *et al.* (2017) respectively.

Despite extensive attempts to gain a larger sample size, achieving the aforementioned samples proved very challenging within the market and timeframe described. Completed survey

responses collated from SurveyMonkey™ across all industries and geographies to which the survey was distributed.

5.2.1.1. Response rate

Though the total polled cannot be accurately listed, given exposure through multiple online and social media platforms, as well as direct communication of in excess of 80 emails, 20 phone calls or 30 direct messages, a total of 109 responses were collected, 96 of which were initially reported as completed by SurveyMonkey™. Given the extent of distribution, the response rate can however be stated as being very low. This completion rate is good, considering that on average, half of respondents do not complete one or more questions (King, Honaker, Joseph, & Scheve, 2001). Remedying incomplete data sets is considered further in the following section. 90 of the respondents passed the initial screening question which prescribed whether respondents qualified to enter the question survey which was “*Are you aware of or associated with a BDA Capability within the organisation being described in this questionnaire?*”, however a total of 80 were sufficiently completed for further application

Akter *et al.* (2016) applied a similar approach to attaining responses through approaching the ‘business analytics’, ‘Big Data analytics’ and ‘IT professionals’ groups, wherein they received a sample twice as large with the only apparent differentiator being those respondents were targeted in the USA, whilst this study targeted primarily South African respondents (Table 5). This may also be an indication of the complexity of the subject matter, which was not only diverse but also not well documented in literature, being assessed relative to the level of penetration of the capability within the South African market, where a majority of respondents were within their first years of being associated to a BDA capability (Table 11). Given this, respondents may have refrained from answering given their lack of exposure, as well as the compounding impact of the Big Data environment being a topic that is not coherently reported, which may make its appreciation for more distantly associated people more difficult. Wamba *et al.* (2017) applied a different approach and used a market research firm that focussed on China, both because of its database of 10 000 Chinese IT managers and because of its reputation which supported collection of a suitable sample.

5.2.2. Total data sample

Despite significant attempts to increase total sample size, a total of 109 respondents completed the survey, whilst only 80 of the initial 109 respondent’s data sets were deemed appropriate for analyses, the results of which are summarised in Table 3. This this presented a

much smaller data set than that which was applied by the reference models for this study, Akter *et al.* (2016) and Wamba *et al.* (2017), who reported total sample sizes of 156 and 297 respondents respectively.

Table 3: Summary of data collected and data imputed for application in analyses

	Number (Total Data Set)	Percentage Total Data	Number Respondents	Percentage of Data Applied in Analyses
Total respondents	109	100.00%	-	-
Total respondents with over 50% questions completed	80	73.39%	80	100.00%
Total respondents' data Imputed	29	26.61%	29	36.25%
Number of Total Potential Answers	7,848	100.00%	5,760	100.00%
Total Number of Answers Collected	5,800	73.90%	5,517	95.78%
Total Number of Answers Missed by respondents	2,048	26.10%	243	4.22%
Total number of Answers Imputed	243	3.10%	243	4.22%

This includes data sets completed through imputation, as described in Section 4.8.4. Therein, of the 5,760 potential answers or data points that could potentially have been captured (80 respondents x 72 questions), 243 data points were missing and were imputed, which is 4.22% of data applied in the study.

Figure 2: Number of Missing Answers per Question and Section of the Questionnaire

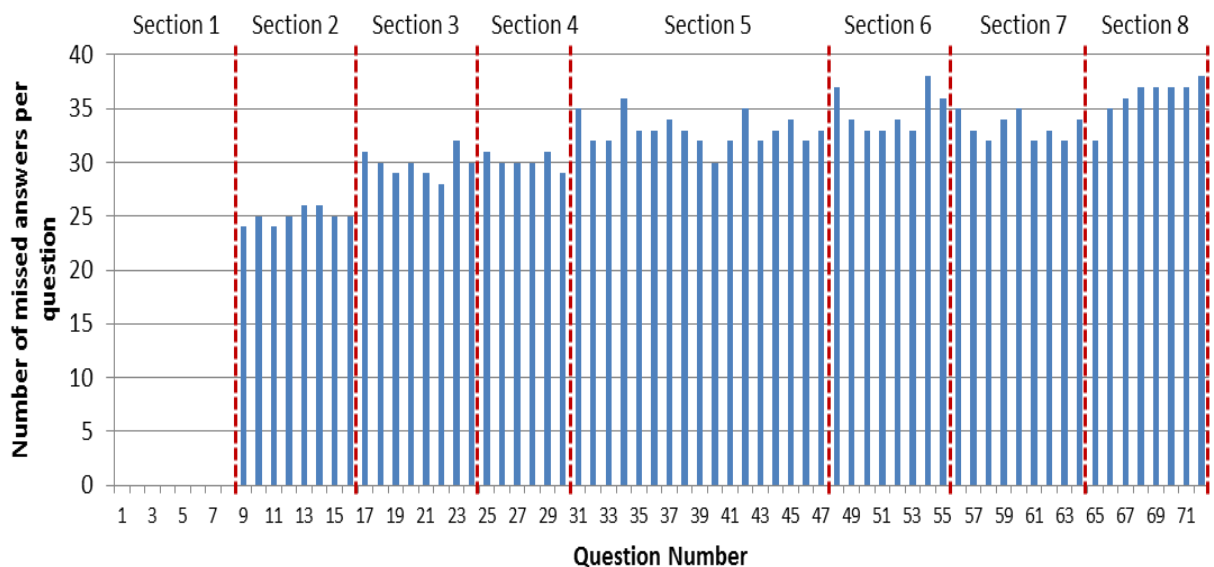


Figure 2 further illustrates the number of missing responses of the sample of 80 respondents per question, as well as illustrating this per section. These sections are also described below for ease of reference through Table 4 which summarises missing and imputed data by section.

Table 4: Summary of missing or imputed data per section of the research survey

Section Detail	Question Range	Missing/Imputed Responses			
		Number	Average	Standard Deviation	Coefficient of Variation
Section 1: Context of Organisation and Respondent	1-8	0	0.00	0.00	-
Section 2: OCBDA	9-16	200	25.00	0.76	3.02%
Section 3: DDDM	17-24	239	29.88	1.25	4.17%
Section 4: FDC and FAC	25-31	185	30.83	2.14	6.93%
Section 5: BDAMC	32-48	563	33.12	1.69	5.11%
Section 6: BDAIF	49-59	375	34.09	1.70	4.99%
Section 7: BDAPE	57-65	297	33.00	1.12	3.39%
Section 8: FPer	66-72	257	36.71	0.95	2.59%

Though no respondents missed answers in section 1, which included descriptive characteristics of respondents, all other sections had an average of 25 or more answers missing per question out of the 80 respondent's data that was used. The increase in average number of missing data from Section 3 (29.88) to Section 8 (36.71) was gradual, with no significant spikes in prevalence.

5.2.3. Normality of data per survey question

As sample size was larger than 30, an appropriately large data set to run a majority of statistical analyses was assumed (Agresti & Franklin, 2007). Given the size of the data set, normality in the distribution data per question was assumed to be appropriate for ongoing analyses, allowing for parametric analyses to be performed (Agresti & Franklin, 2007). As descriptive and analytical assessments did not directly require direct comparison of individual questions from the survey (such t-tests to assess differences) direct normality in distribution of data of individual questions was furthermore not influential. Representation of the normality of each question is however described through skewness as a descriptive assessment of each question in Appendix 4. Therein, the average skewness statistic and standard error were -0.277 and 0.269 respectively, illustrating that most questions were largely normally distributed.

5.2.4. Descriptive characteristic of respondents and their organisations

A total of 8 biographic variables were used to profile the respondents, towards informing and categorising relative traits of their responses and how this might influence subsequent analyses. These included an initial screening question which interrogated whether respondents were aware of or associated with a BDA capability in the organisation being described. As all

80 respondents answered affirmatively to this, in the sample applied, not further statistical analyses were performed on this. Further measures assessed included the principal industry of respondent's organisations, current job level, personal experience with the field of BDA, type of department of work relative to BDA, geographic location of work, duration that the organisations have actively pursued or applied BDA in their businesses, and approximate total number of employees within organisations. In each of the charts presented in this section, the labels include first the absolute number of respondents, and the percentage of total respondents that represents in each case. Moving from general to more personal characteristics, the geographic location of respondents is considered first, followed by industry, organisational and finally personal characteristics. Descriptive statistics of each question are summarised through Table 29 in Appendix 4, and which is referenced in relation to each of the figures in this section.

5.2.4.1. Country in which respondents work

As is illustrated in Table 5, 96.25% of respondents whose data were applied in the analyses work in South Africa, which describes a near homogenous sample.

Table 5: Country in which respondents work

Measure	Count	Percent of total
South Africa	77	96.25%
Luxembourg	1	1.25%
United Arab Emirates	1	1.25%
Ireland	1	1.25%
Total	80	100.00%

Assessment of the industries represented is relevant within the study as not all industries are currently capable of, or rather positioned to easily collect sufficient data to be classified as Big Data, nor to apply BDA within their businesses. The results of industry assessment are provided through Table 6 below.

5.2.4.2. Industries represented in the survey

Unsurprisingly for the South African market, telecommunications, technology, internet and electronics, as well as finance and financial services were the two most prevalent sources of respondents with 35% and 29% of the respondent base respectively (Table 6). These are due to the large data streams which one would collect in these industries through the very nature of the operations of business, where telecommunications would be based largely on call data

records (CDRs) and electronic banking transactions, both of which are made by individual consumers on these organisations infrastructure. Though Wamba *et al.* (2017) reported a very similar weighting to telecommunications, technology, internet and electronics through their definition ‘*Information and communication*’ the remainder of their sample was more evenly weighted, possible as a factor of sample size which was 225. These were followed by respondents from the insurance market (10%), which are relatively high data yielding through the depth of data collected on customers to appropriately determine risk profiles. The transportation industry followed this with 6% of respondents. The major data flowing through these businesses most likely being in the form of telemetry, or tracking data reporting aspects such as vehicle location, breaking and engine usage parameters.

Table 6: Industries represented by survey

Measure	Count	Percent of total
Food & Beverages	1	1.25%
Finance and Financial Services	23	28.75%
Transportation & Delivery	5	6.25%
Telecommunications, Technology, Internet & Electronics	28	35.00%
Insurance	8	10.00%
Manufacturing	2	2.50%
Retail & Consumer Durables	6	7.50%
Utilities, Energy, And Extraction	2	2.50%
Agriculture	2	2.50%
Automotive	1	1.25%
Non-profit	1	1.25%
Education	1	1.25%
Total	80	100.00%

5.2.4.3. Relative organisation size

Each of the industries described are predominated by large organisations, which is reflected in the headcount of employees as illustrated in Table 7. Over half (54%) of the respondents came from organisations with 1000 or more employees, whilst a further 21% come from organisations with over 500 employees, which amounts to $\frac{3}{4}$ of the sample combines. This is a feature which was expected given the volume of data that is required to require a Big Data platform, as well as the typically high cost of maintaining such an infrastructure and employing appropriately skilled personnel.

Table 7: Number of employees within respondents organisations

Measure	Count	Percent of total
1 – 99	12	15.00%
100 – 499	6	7.50%
500 – 999	17	21.25%
1,000 or more	43	53.75%
Don't know	2	2.50%
Total	80	100.00%

5.2.4.4. Organisation's tenure in BDA

The final descriptive measure of the organisations themselves describes the time that these organisations have actively pursued BDA. This is significant firstly as it describes the potential for the maturity of the base BDA capability. Secondly, and as critically, it also alludes to a potential for the level of the maturity of the implementation in a business, and the potential for respondents to appreciate the capability itself, and to objectively assess what its level of application might be relative to what a fully or well implemented and mature BDA capability may be. The results of this are illustrated through Table 8.

Therein, 52% of organisations were reported to be within the first two years of their BDA deployments within business, corresponding to the perceived infancy of the capability within firms in South Africa. A further 26% of organisations have reportedly actively pursued a BDA capability for up to 6 years, corresponding with the rapid increase in popularity of Big Data as a capability from 2011 as reported by Lee (2017). A further 6% reported having actively pursued BDA for 6-8 years, whilst 16% reported that their organisations had pursued BDA for more than 8 years, suggesting their active involvement since as 2008 or earlier. This is interesting considering that the tools for Big Data platforms were only in development from 2007, with general notoriety only having been gained from 2008 when *Nature* published a special issue on Big Data (Chen *et al.*, 2014; Lee, 2017).

Table 8: Years for which respondents organisations have actively pursued BDA

Measure	Count	Percent of total
0 – 2	41	51.25%
2 – 4	8	10.00%
4 – 6	13	16.25%
6 – 8	5	6.25%
8+	13	16.25%
Total	80	100.00%

5.2.4.5. Respondent's seniority in business

Moving on to descriptions of the respondents themselves, the following figures consider aspects which relate to the association or respondents themselves with the BDA environment. As is illustrated in Table 9, which describes the relative seniority of positions held by respondents in their respective organisations, only 2% reported themselves as being entry level, with a relatively even split across the remaining four levels described of 23%, 30%, 35% and 20% respectively for intermediate, middle management, senior management, and finally owner, executive or C-suite level. This suggests a relatively senior sample, which was necessary given the required exposure that respondents would require to complete the survey.

Table 9: Current job level of respondents

Measure	Count	Percent of total
Entry Level	2	2.50%
Intermediate	18	22.50%
Middle Management	24	30.00%
Senior Management	20	25.00%
Owner/Executive/C-Level	16	20.00%
Total	80	100.00%

5.2.4.6. Association to the BDA environment

In addition to the seniority of respondents within organisations, their association or type of role relative to the BDA environment was also considered. This measure was intended to describe the relative position from which respondents considered the BDA capability, and was intended to describe context toward the expressed perceptions these respondents may have had toward organisational, technical or analytical competencies of the organisation as all of there were relevant to the context of the study.

Table 10: Main association with the BDA capability

Measure	Count	Percent of total
User of analytics within business	27	33.75%
Data analyst (Direct processor of data)	24	30.00%
Big Data Management (Driving application of resources)	11	13.75%
IT Systems or Infrastructure (Data technology environment)	15	18.75%
Consultant	3	3.75%
Total	80	100.00%

As presented in Table 10, users of analytics within business comprised a third of the sample (33%), which is significant as these respondents were assumed to have a more neutral

perception of a BDA capability as well as its relative impact on the business than the 24% of respondents who were directly associated with analytics. Those respondents, along with the 19% IT systems or infrastructure would however more likely be able to present a deeper technical understanding of the environment, and express a perceived impact from this perspective. At nearly one third of the sample each, this provides a balance of these perspectives.

The final one third of the sample comprised of respondents associated with upper management or advisory to BDA capabilities (note 3 respondents with related terms described in the 'other' category were included here as they were closely associated), and thus would provide a view of the capability within context of the business. Given the dispersion of respondents, the sample was considered well balanced from this perspective.

5.2.4.7. Personal experience in BDA

The final aspect considered was that of respondent tenure relative to the field of BDA, as a measure of the level of direct experience that respondents had with respect to the environment being considered. As shown in Table 11, 36% of respondents reported between 0-2 years' experience with the BDA environment, whilst 31% reported 2-5 years' experience, which combined is nearly 2/3 of the total, a result which is similar to the extent of organisations commitment to BDA in Table 8. Furthermore, similarly to that result, a further 25 reported between 5-10 years' experience in BDA, with 8% suggesting they had more than 10 years' experience in BDA.

Table 11: Years of work experience in the field of BDA

Measure	Count	Percent of total
0 – 2	29	36.25%
2 – 5	25	31.25%
5 – 10	20	25.00%
10 – 15	3	3.75%
15+	3	3.75%
Total	80	100.00%

5.3. Assessment of Survey Sections as Constructs

Section 1 of the survey, which included questions which related to context of respondents themselves and their organisations (Section 5.2.4), was not included in the following as it

pertains to descriptive characteristics rather than constructs assessed by Likert scale as ordinal variables.

5.3.1. Internal reliability of data sets (Cronbach's alpha)

Cronbach's alpha is a commonly applied measure of internal consistency, or reliability of measures or scales, and is often used where multiple Likert questions are present in a survey, as in this study (Bland & Altman, 1997; Kline, 2016). Assessment constructs was performed as distinct sections of the survey to determine internal reliability prior to further application in the study. The results of those assessments are summarised in Table 12. Therein it is evident that all sections assessed returned Cronbach alpha scores in excess of 0.7, thus all constructs were reliable (Kline, 2016).

It should be noted that initial assessments of Sections 3 (DDDM) and 4 (FDC) both yielded Cronbach's alpha scores lower than 0.7, at 0.449 and 0.622 respectively. Each section was amended by deletion of one question at a time, as described in Appendix 4 (Section 0, and illustrated in Table 30), with a total of two questions deleted in DDDM by deletion of 1 question deleted from FDC. Consequentially, all constructs had $\alpha > 0.7$, thus construct could be applied in this form on subsequent analyses.

Table 12: Summarised reliability statistics of internal sample data (Cronbach's alpha)

Scale	Sample Size	Number of Items prior to Cronbach alpha	Number of Items post Cronbach alpha	Cronbach Alpha	Decision
Section 2: OCBDA	80	8	8	0.838	Proceed
Section 3: DDDM	80	8	6	0.772	Proceed
Section 4: FDC and FAC	80	7	6	0.839	Proceed
Section 5: BDAMC	80	17	17	0.915	Proceed
Section 6: BDAIF	80	8	8	0.724	Proceed
Section 7: BDAPE	80	9	9	0.887	Proceed
Section 8: FPer	80	7	7	0.933	Proceed

5.4. Factor Analysis of constructs

5.4.1. Exploratory Factor Analysis

Exploratory factor analyses (EFA) was performed on each of the survey constructs, which was designated as constructs in the survey. Suitability of the data to have EFA analyses performed was addressed through sections 0 in Appendix 4, along with Section 5.3 (internal reliability)

and Section 4.9.3.3 (internal validity) respectively, all of which ensured assumptions were met for the constructs assessed. The results of EFA by scale are summarised through Table 13 below, and are discussed hereafter.

Theoretical constructs assessed in relation to the base model were compiled by the researcher, and comprised of a total of 8, 8 and 7 questions for OCBDA, DDDM and FDC respectively, whilst each had 80 responses. In each case, the scales and questions applied were extracted from literature relating to the specific components being assessed, however entire constructs were not applied in this study. This was due both to a restriction in the number of question that could be asked due to sample length, and due to sub components of questionnaires being of direct interest in this study. Due to this, direct comparison of construct means from literature of each construct was not possible in relation to these data in each case, constructs were however well established.

Table 13: Summary results from Exploratory Factor Analysis to identify factors from survey constructs

Scale	Section 2: OCBDA	Section 3: DDDM	Section 4: FDC and FAC	Section 5: BDAMC	Section 6: BDAIF	Section 7: BDAPE	Section 8: FPer
Sample Size	80	80	80	80	80	80	80
Number of Items	8	8	7	17	8	9	7
Cronbach Alpha	0.838	0.772	0.839	0.915	0.724	0.887	0.933
KMO measure of sampling adequacy	0.845	0.790	0.815	0.863	0.795	0.813	0.897
Bartlett's test of sphericity p	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Percentage of variance extracted	61.38	48.651	73.94	68.361	62.94	53.344	72.599
Number of factors extracted (Eigenvalue ≥ 1)	2	1	2	4	2	2	1
Validity	All items $p \leq 0.01$	All items $p \leq 0.01$	All items $p \leq 0.01$	All items $p \leq 0.01$	All items $p \leq 0.01$	All items $p \leq 0.01$	All items $p \leq 0.01$
Correlation p value	≤ 0.01	≤ 0.01	≤ 0.01	≤ 0.01	≤ 0.01	≤ 0.01	≤ 0.01
Decision	Proceed	Proceed	Proceed	Proceed	Proceed	Proceed	Proceed

Within OCBDA, questions were extracted from instruments with extensive application in research which were designed to assess entrepreneurship, or entrepreneurship within a corporate setting (Hill, 2003; Kuratko *et al.*, 1990), which was positioned as an appropriate construct base to assess organisational entrepreneurship towards a capability such as BDA. Kuratko *et al.* (1990) reported Cronbach alpha scores per question that ranged from 0.68 to 0.75 and correlated item-total correlations that ranged from 0.39 to 0.60. The current study combined shortened scales from the two sources, and found a comparatively higher

Cronbach's alpha of 0.838. The scale was valid, with correlation coefficients > 0.47 and $p < 0.00$ (Table 32). The overall KMO sampling adequacy was 'meritorious' at 0.845 (Kaiser, 1974), along with significant score in the Bartlett's test for sphericity at $p = 0.00$ indicating that the data was factorisable. Therefore the data was deemed suitable for PCA analysis (Hair *et al.*, 2010; B. Williams *et al.*, 1996; Zikmund *et al.*, 2012). EFA revealed two components with eigenvalues greater than 1 and which explained cumulative variance of 61.38% (Table 34), a result also validated by visual inspection of scree plot (Figure 5) (Costello & Osborne, 2005). The contents of the factors were interpreted through Varimax orthogonal rotation (Table 35), with the mean of resulting question groupings applied in subsequent analyses.

5.4.2. Confirmatory Factor Analysis

CFA was used to verify the convergent and discriminant validity of the measurement model, and was done so at a first order level for the respective constructs. The results of CFA are summarised in Table 14, and presented in detail in Section 9.4.6. Therein, an acceptable model fit was identified with root mean square error of approximation (RMSEA = 0.063), a comparative fit index (CFI = 0.972) and an incremental fit index (IFI = 0.974), all of which were the criteria for a good model fit (Hu & Bentler, 1999; Jackson, Gillaspay, & Purc-Stephenson, 2009). Factor loading ranged from $\lambda = 0.00$ to 0.85. Along with a significant t-statistic (1.311), this suggests firstly that factors loaded more on their own than on other constructs and that they all share a substantial amount of variance with their own items (Akter *et al.*, 2016). This provides further evidence of validity of construct as CFA provides a "good tool for assessing construct validity because it provides a test of how well the researcher's "theory" about the factor structure fits the actual observations" (Zikmund *et al.*, 2012, p593).

Table 14: Summary results from Confirmatory Factor Analysis per construct

Scale	IFI	CFI	RMSEA	λ	t-statistic
OCBDA	0.974	0.972	0.063	2 of 8 = $> .50$	1.311
DDDM	0.992	0.992	0.038	5 of 6 = $> .50$	1.114
FDC	0.995	0.995	0.039	2 of 6 = $> .50$	1.119
BDAMC	0.876	0.871	0.100	all $> .50$	1.792
BDAIC	0.880	0.875	0.143	all $> .50$	2.624
BDAPE	0.866	0.862	0.158	all $> .50$	2.976
Fper	0.972	0.971	0.118	all $> .50$	2.099
BDAC	0.941	0.939	0.115	7 of 8 = $> .50$	2.036

DDDM questions were extracted from an assessment tool designed by the *Center for Evidence-Based Management* (CEBMA, 2013). Given that truncated sections of the questionnaire were extracted, and that it was not explicitly validated therein, it was necessary to ensure both reliability and validity of the construct prior to EFA as an exploratory measure, as well as PCA posts that application such that it could be directly applied in the BDA-FPer model. The current study found a Cronbach's alpha of 0.772 confirming its reliability. Validity of the construct was confirmed through correlation coefficients all > 0.41 and $p < 0.00$ (Table 32). The overall KMO sampling adequacy was 'middling' at 0.772 (Kaiser, 1974). Bartlett's test for sphericity was significant at $p = 0.00$ showing that data was factorisable, and suitable for PCA analysis (Hair *et al.*, 2010; B. Williams *et al.*, 1996; Zikmund *et al.*, 2012). EFA revealed a single factor loading explaining 48.651% of variance, as assessed through eigenvalues and scree plot (Costello & Osborne, 2005). Factor components were interpreted through Varimax orthogonal rotation, with the mean of resulting question groupings applied in subsequent analyses. The results of DDDM CFA (Table 14 and Section 9.4.6) illustrate an acceptable model fit with RMSEA = 0.038, CFI = 0.992 and IFI = 0.992, all of which were above the critical values for a good model fit (Hu & Bentler, 1999; Jackson *et al.*, 2009). Factor loading was large in each case ($\lambda > 0.5$), along with a significant t-statistic = 1.114, again confirming validity of the construct (Hair *et al.*, 2010).

Similarly to OCBDA, Firm Dynamic Capability (FDC) and Firm Adsorptive Capacity (FAC) was comprised of questions compiled from across multiple sources (Elbashir *et al.*, 2011; Ravichandran & Lertwongsatien, 2005; Sher & Lee, 2004; N. Wang *et al.*, 2012). With a Cronbach's alpha of 0.839 was reported thus reliability was confirmed with a higher result than Ravichandran and Lertwongsatien (2005) ($\alpha = 0.65$) for a similar construct composition. Validity of the construct was confirmed through correlation coefficients all > 0.55 and $p < 0.00$ (Table 32). The overall KMO sampling adequacy was 'meritorious' at 0.815 (Kaiser, 1974). Bartlett's test for sphericity was significant at $p = 0.00$ again showing that data was factorisable, and suitable for PCA analysis (Hair *et al.*, 2010; B. Williams *et al.*, 1996; Zikmund *et al.*, 2012). Analysis of eigenvalues and scree plot (Costello & Osborne, 2005) revealed two factors that described a cumulative variance of 73.94%, higher than that reported by Sher and Lee (2004) at 69.592% for a similar construct composition. CFA analyses presented acceptable model fit with RMSEA = 0.039, CFI = 0.995 and IFI = 0.995, again within values that describe a good model fit (Hu & Bentler, 1999; Jackson *et al.*, 2009). Factor loading ranged from $\lambda = 0.00$

to $\lambda = 0.73$, and had t-statistic = 1.119, providing further evidence of construct validity (Zikmund *et al.*, 2012, p593).

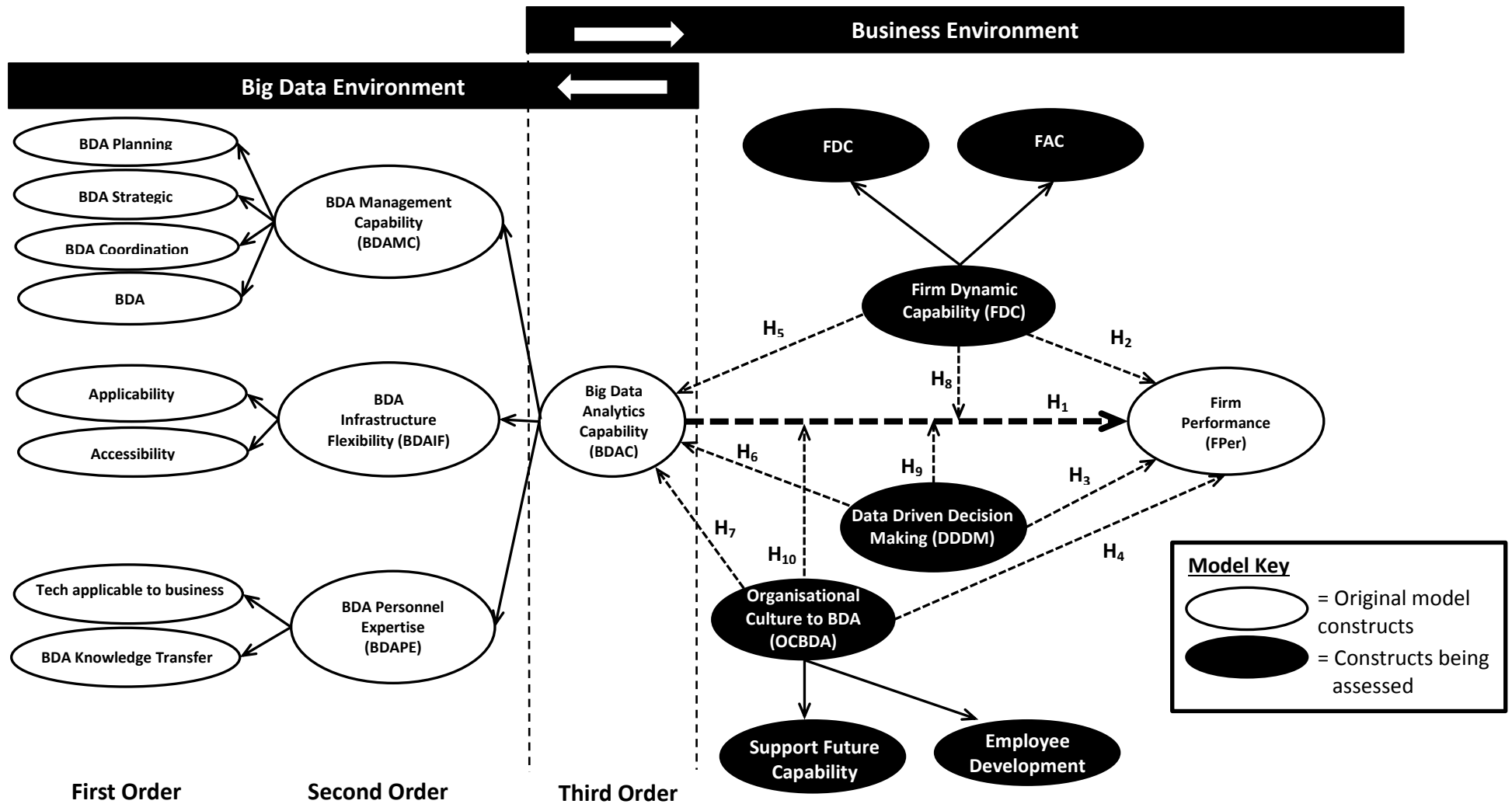
As BDAMC, BDAIF, BDAPE and FPer were based on the model described by Akter *et al.* (2016) and Wamba *et al.* (2017), thus are describes together, and in relation to those studies. In all cases, reliability of constructs was confirmed with $0.724 > \alpha < 0.933$. Validity was also confirmed with Pearson's correlations scores ranging from 0.462 to 0.926, all with $p \leq 0.000$. KMO sampling adequacy ranged from the upper ranges of 'middling' to 'meritorious' $0.795 > KMO < 0.897$ (Kaiser, 1974), whilst Bartlett's test for sphericity were all significant at $p = 0.00$, thus constructs were all factorisable, and could be forwarded to PCA analysis (Hair *et al.*, 2010; B. Williams *et al.*, 1996; Zikmund *et al.*, 2012). Analysis of eigenvalues described 4, 2, 2 and 1 construct for BDAMC, BDAIF, BDAPE and FPer factors respectively, each describing a cumulative variance of 68.36%, 62.94%, 53.34% and 72.59% the same constructs respectively, all of which are closely aligned with by Akter *et al.* (2016) and Wamba *et al.* (2017). Although both those authors proceeded with structured equation modelling (SEM), the remainder of the factors and results are aligned with regard to composition, though differences in factor composition and loading were present. This was expected to be a result of sample size, differences in methods applied, as well as variations in survey composition due to constraints on survey length imposed on the current study.

For BDAMC, BDAIF, BDAPE and FPer factors respectively, CFA analyses presented acceptable model fits with RMSEA = 0.100, 0.143, 0.158, 0.118 and 0.115, CFI = 0.871, 0.875, 0.862, 0.971 and 0.939 and IFI = 0.876, 0.880, 0.866, 0.972 and 0.941. The RMSEA values were higher than the guidelines described by (Hu & Bentler, 1999), and suggests that total model fit for these parameters was not good. CFI and IFI values were, in cases also lower than the close to or greater than 0.95 value reported by Hu and Bentler (1999). The levels presented by Hu and Bentler (1999) have however been contested as being too conservative for many types of models (Brown & Moore, 2012; Marsh, Hau, & Wen, 2004). Given this and the fact that though low, the results were not extremely different, the CFI and IFI values were considered to be applicable. Furthermore, as the three measures are reported together to provide a more holistic description of model fit (Brown & Moore, 2012), the presence of two reasonable fit results and one poor result allow for continued application, but with a limitation on the extent to which results could be extrapolated, due to concern surrounding the goodness of fit. Factor loading ranged from $\lambda = 0.5$ to $\lambda = 0.95$, which along with all t-statistic > 1.792 , illustrating

validity (Zikmund *et al.*, 2012, p593). These results compared favourably with the reference studies of who also applied CFA (Akter *et al.*, 2016; Wamba *et al.*, 2017).

Similarly, a model of the third order construct, BDAC, was also built in as validate its loading relative to literature (Figure 13). As illustrated through Table 14, data compared closely with other constructs of the model, and also validated it. Importantly, as these results compared closely with Akter *et al.* (2016) and Wamba *et al.* (2017), it can be concluded that the base model was successfully reproduced, and that the added constructs could thus be related similarly.

Figure 3: Revised research model applied in the study



5.5. Construct Descriptive Statistics

Given the results of both EFA and CFA, the constructs described above were applied in the testing of the hypotheses presented in Chapter 3. As the constructs had been factored however, their resulting traits were assessed with descriptive stats to characterise them as their attributes had the potential to influence analyses performed. An assessment of descriptive statistics for each of the constructs is summarised in Table 15. Therein mean and standard deviation for each factor is presented, the majority of which display a mean in the region of 4.6, suggesting skewness of the data set with tendency toward positive answers by respondents across all responses. This is further supported by the spread of max and min variables, where though the maximum of 7 (highly positive answers) is reached, the minimum of 1 (highly negative answers) is not, suggesting an inclination to agree with the position of the research by respondents, and a resulting likelihood that data would correlate. The coefficient of variation of the data ranged from 16% up to 33%, describing relatively high variance in data. No significant outliers were visible as assessed through scatter plots of each factor compared to FPer, which also confirmed linearity in the data.

Table 15: Descriptive statistics of factors extracted from EFA and CFA

Descriptive Statistics										
1 st order Constructs	2 nd order Constructs	N Statistic	Min. Statistic	Max. Statistic	Mean Statistic	Std. Deviation Statistic	Skewness		Kurtosis	
Support Future Capability Employee Dev.	OCBDA	80	2.750	7.000	5.572	0.949	-0.923	0.269	0.662	0.532
		80	1.500	7.000	4.638	1.255	-0.239	0.269	-0.570	0.532
DDDM	DDDM	80	2.500	6.833	4.569	0.927	0.080	0.269	-0.231	0.532
FAC FDC	FDC and FAC	80	1.500	7.000	4.750	1.425	-0.644	0.269	-0.546	0.532
		80	1.000	6.750	4.131	1.351	-0.280	0.269	-0.828	0.532
Strategic Coordination Planning Implementation	BDAMC	80	2.000	7.000	4.625	0.989	-0.106	0.269	-0.217	0.532
		80	1.400	6.800	4.173	1.083	0.080	0.269	-0.063	0.532
		80	1.714	6.571	4.548	0.911	-0.684	0.269	0.648	0.532
		80	2.000	6.667	4.408	1.095	-0.209	0.269	-0.135	0.532
Applicability Accessibility	BDAIF	80	1.500	6.500	4.152	1.154	-0.250	0.269	-0.292	0.532
		80	2.000	6.500	4.213	0.860	0.230	0.269	0.082	0.532
Tech applicable to business BDA Knowledge Transfer	BDAPE	80	2.800	7.000	5.308	0.964	-0.786	0.269	0.424	0.532
		80	2.250	7.000	5.019	1.035	-0.570	0.269	0.502	0.532
Firm Performance	FPer	80	1.000	6.857	4.455	1.147	-0.543	0.269	0.733	0.532

Given the skewness present in the data, A Shapiro-Wilk's test (Table 16) was performed on factors to assess normality of data prior to correlation and regression analyses (Razali & Wah, 2011; Shapiro & Wilk, 1965).

Table 16: Assessment of factor normality through Shapiro-Wilk testing

Tests of Normality	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	df	<i>p</i>	Statistic	df	<i>p</i>
Support Future Capability	0.157	80	0.000	0.930	80	0.000
Employee Development	0.089	80	0.187	0.981	80	0.289
DDDM	0.083	80	0.200*	0.990	80	0.787
FAC	0.157	80	0.000	0.923	80	0.000
FDC	0.152	80	0.000	0.965	80	0.027
Strategic	0.100	80	0.045	0.977	80	0.152
Coordination	0.076	80	0.200*	0.989	80	0.725
Planning	0.123	80	0.005	0.960	80	0.014
Implementation	0.098	80	0.057	0.978	80	0.172
Applicability	0.105	80	0.029	0.980	80	0.247
Accessibility	0.135	80	0.001	0.967	80	0.036
Tech applicable to business	0.124	80	0.004	0.943	80	0.001
BDA Knowledge Transfer	0.118	80	0.008	0.966	80	0.031
Firm Performance	0.158	80	0.000	0.964	80	0.025

The null-hypothesis is rejected where $p < 0.05$ showing, resulting in the presumption of a statistically sig diff between construct and the normal distribution, and thus that the sample is not normally distributed. As described in Table 16, and through visual inspection of their histograms, 8 of the 14 constructs were not normally distributed, a result also confirmed by Kolmogorov-Smirnov analyses. As a result, non-parametric correlations (Spearman's Rho) were also run on to assess correlation of constructs. Although normality was violated, Maxwell and Delaney (2004) suggest that analyses can continue without transformations.

5.6. Assessment of Research Hypotheses

5.6.1. Correlation of research constructs

The following section considers research questions one and two of the study, as it assesses the direct relationship of constructs with FPer. In so doing, this section establishes the base relationship between the factors, and significantly describes whether the factors have an impact on the output of FPer.

5.6.1.1. Research question 1

The first research question considered whether there was a relationship between Big Data Analytics Capability (BDAC) and Firm Performance (FPer), and did so through H_1 which posited the presence of a positive relationship between the two, with FPer as the dependent variable. As described in Section 5.4, and presented in Figure 13, CFA confirmed the appropriateness of the third order factor, BDAC as a cumulative result of BDAMC, BDAIF and BDAPE, thereby allowing for it to be applied directly as a construct. The results of Pearson's and Spearman's correlations are summarised in Table 17 for all factors assessed, are presented in relation to FPer. A correlation matrix of each factor is included in Table 36 for reference.

Table 17: Correlation coefficients of constructs with Firm Performance

Firm Performance	Pearson Correlation Coefficient	p (2-tailed)	N	Spearman's Rho Correlation Coefficient	p (2-tailed)	N
Support Future Capability	.346**	0.002	80	.391**	0.000	80
Employee Dev	.388**	0.000	80	.397**	0.000	80
DDDM	.583**	0.000	80	.524**	0.000	80
FAC	.309**	0.005	80	.318**	0.004	80
FDC	.666**	0.000	80	.636**	0.000	80
Strategic	.357**	0.001	80	.436**	0.000	80
Coordination	.622**	0.000	80	.621**	0.000	80
Planning	.632**	0.000	80	.648**	0.000	80
Implementation	.517**	0.000	80	.553**	0.000	80
Applicability	.529**	0.000	80	.516**	0.000	80
Accessibility	-0.032	0.775	80	0.051	0.654	80
Tech applicable	.265*	0.018	80	.341**	0.002	80
Knowledge Transfer	0.215	0.055	80	.305**	0.006	80
BDAC	.632**	0.000	80	.610**	0.000	80

** . Correlation is significant at the 0.01 level (2-tailed).

* . Correlation is significant at the 0.05 level (2-tailed).

As presented in in Table 17, BDAC correlates strongly with FPer with a Pearson's Correlation of 0.632 and $p = 0.00$. Given this the null-hypothesis is rejected, and H_1 is confirmed, showing that in the presence of an increased BDA capability, FPer too increases, thus a BDA Capability has a positive impact on FPer, a result which concurs with literature (Akter *et al.*, 2016; Wamba *et al.*, 2017, 2015). This relationship is further validated as each of the components of BDAC were also assessed, which included BDAMC, BDAIF and BDAPE. The first order factors assessed (first order construct loading to second order constructs is summarised through Table 15) in relation to FPer all presented with significant positive correlations in their own capacity

(Table 17), other than 'Accessibility'. Though 'Knowledge Transfer' also showed insignificance in Pearson's correlation, it was found to be non-normally distributed (Table 15), and thus when non-parametric assessment run, it was significant. These results thus confirm that BDAC does present a strong positive correlation with FPer, both at a third order level, and at a first order level as measures by its respective components.

5.6.1.2. Research question 2

Research question 2 assessed the relationship between the three organisational constructs added to the model in this study (FDC, DDDM and OCBDA) with FPer. Though prior studies have described links between organisational capabilities and FPer (Akter *et al.*, 2016; Chen & Zhang, 2014; Wamba *et al.*, 2017), the constructs in this form have not previously been tested. This study assessed these relationships through H₂, H₃ and H₄.

As presented in Table 17, strong correlations were found for each of the research constructs. As illustrated in Figure 3 and, FDC and OCBDA both factorised to two components. Within FDC (H₂), Firm adsorptive capacity (FAC) resulted in a correlation coefficient of 0.318 ($p = 0.040$) confirming its positive correlation with FPer. Firm dynamic capability presented a stronger correlation at 0.636 ($p = 0.000$), describing a strong correlation, both of which rejected the H₂ null-hypothesis. Similarly, OCBDA was comprised on 'Support Future capability' (.391, $p = 0.000$) and 'Employee Development' (.397, $p = 0.000$), both of which were thereby significant correlations with FPer, rejecting null-hypothesis of H₄. DDDM retained its single component structure, and resulted in a strong significant correlation (.524, $p = 0.000$), rejecting null-hypothesis of H₃, and confirming correlation of the third and final organisational construct assessed.

5.6.1.3. Research question 3

Research question 3 similarly assessed the relationship between the three organisational constructs added to the model in this study (FDC, DDDM and OCBDA) with BDAC to confirm whether there was a direct relationship between the organisation and the BDA capability itself. This study assessed these relationships through H₅, H₆ and H₇. As presented in Table 18, strong correlations ($p < 0.001$) were found for all construct with relation to BDAC, and furthermore did so with higher correlation coefficients than was seen in Table 17 for FPer. This therefore results in the null-hypotheses of through H₅, H₆ and H₇ all being rejected. Accordingly, FDC, DDDM and OCBDA (as their respective factors) correlate strongly with BDAC. Thus the

organisations position toward BDA does have a significant impact on the functioning or effectiveness of that capability within that organisation.

Table 18: Correlation coefficients of constructs with BDA capability

BDA Capability	Pearson			Spearman's Rho		
	Correlation Coefficient	ρ (2-tailed)	N	Correlation Coefficient	ρ (2-tailed)	N
Support Future Capability	.553**	0.000	80	.528**	0.000	80
Employee Dev	.588**	0.000	80	.576**	0.000	80
DDDM	.657**	0.000	80	.680**	0.000	80
FAC	.306**	0.006	80	.344**	0.002	80
FDC	.652**	0.000	80	.661**	0.000	80
Strategic	.613**	0.000	80	.607**	0.000	80
Coordination	.773**	0.000	80	.800**	0.000	80
Planning	.831**	0.000	80	.844**	0.000	80
Implementation	.831**	0.000	80	.823**	0.000	80
Applicability	.783**	0.000	80	.787**	0.000	80
Accessibility	.340**	0.002	80	.334**	0.002	80
Tech applicable	.638**	0.000	80	.656**	0.000	80
Knowledge Transfer	.658**	0.000	80	.659**	0.000	80
FPer	.632**	0.000	80	.610**	0.000	80

** . Correlation is significant at the 0.01 level (2-tailed).

Unsurprisingly, the factors with the highest correlation coefficient are those that pertain to direct capabilities within the BDA environment, which are those which pertain to BDAMC, BDAIF and BDAPE (Figure 3). Though Accessibility is from within that environment, it had a comparatively low coefficient (0.334) which aligns with the descriptor of that factor which pertains to the ability of the organisation to access the BDA functions outputs. The factors of the three respective added construct (Figure 3) from the business environment each also returned large correlations, with OCBDA (Support future capability = 0.528, Employee development = 0.576) describing acceptance and willingness of the organisation to embrace such a capability. DDDM similarly had a large coefficient (0.680) describing the prominence of the outputs of a BDA capability toward driving business decisions. The FDC construct showed strong correlation for the FDC factor (0.661), but a comparatively lower coefficient for FAC (0.344), potentially illustrating that the ability of the organisation to embrace the capability is more significant at the early stages of BDA take on (52% of respondents organisations were within the first 2 years on BDA take on, Table 8). Significantly however, these results describe the strong influence which the organisation has on whether the BDA capability will function effectively.

5.6.2. Hierarchical multiple regression of research constructs

Though research questions 1 and 2 describe the presence of individual impact on BDAC and FPer respectively, the following research question assessed the relative cumulative impact of the factors such that their weighting of influence within the environment could be assessed.

5.6.2.1. Research Question 4

Research question 4 was the primary research question of the study, and assessed the combined impacts of FDC, DDDM and OCBDA as organisational capabilities or attributes on the BDAC – FPer relationship, specifically assessing whether the relationship was moderated by the organisational capabilities. These are assessed through H_8 , H_9 and H_{10} , with the respective HMR outputs presented through Table 19, wherein FDC, OCBDA, DDDM and BDAC respectively were added to a first order model that was built in their respective absence, such that their relative impact on the factors applied were run in level model two respectively in order to characterise their respective impact on the environment.

Given the nature of these constructs, FDC, DDDM and OCBDA are capabilities within an organisation which were anticipated to impact on FPer through their impact the functioning of the BDAC function. Hierarchical regression revealed all models were significant (ANOVA $p = 0.000$) and thus could be applied in assessing relative influence to FPer through variance explained. The ΔR^2 in each Model 1 ($p < 0.05$) describes the variance explained in the absence of the variable considered added in Model 2. In the absence if each of the following respectively (each was present in the other models unless stipulated it was loaded in Model 2); OCBDA = 59.9%, DDDM 62.0%, FDC 56.1%, and BDAC 51.4%. The individual contribution of these factors as identified by ΔR^2 in each Model 2 was OCBDA = 4.43%, DDDM 2.31%, FDC 8.20%, and the largest contributor being BDAC at 12.84%. Total model variability explained after adding constructs in Model 2 respectively was reported by R^2 as being OCBDA = 64.3%, DDDM 64.3%, FDC 64.2%, and BDAC 64.2% respectively, showing relative congruence in model output. Unsurprisingly, the factors with the lowest impact in Model 2 had the highest variance explained in their absence in Model 1.

Table 19: Summary of hierarchical regression analysis of factors (FDC, OCBDA and DDDM) impacting on the BDAC - FPer relationship

	Scale	Model			ANOVA <i>p</i>	Coefficient		
		<i>R</i>	<i>R</i> ²	ΔR^2		β	<i>t</i>	<i>p</i>
OCBDA	Model 1	0.774	0.599	59.9%	0.000			
	FDC					0.371	2.880	0.005
	Strategic					-0.193	-1.786	0.079
	Accessibility					-0.181	-2.158	0.034
	Model 2 (OCBDA added)	0.802	0.643	4.43%	0.000			
	DDDM					0.257	2.066	0.043
	FDC					0.511	3.846	0.000
	Coordination					0.317	2.136	0.036
	Accessibility					-0.170	-2.107	0.039
DDDM	Model 1	0.787	0.620	62.0%	0.000			
	Employee Development					-0.253	-1.982	0.052
	FDC					0.544	4.025	0.000
	Coordination					0.374	2.506	0.015
	Accessibility					-0.157	-1.911	0.060
	Model 2 (DDDM added)	0.802	0.643	2.31%	0.000			
	Employee Development					-0.339	-2.583	0.012
	FDC					0.511	3.846	0.000
	Coordination					0.317	2.136	0.036
FDC	Model 1	0.749	0.561	56.1%	0.000			
	DDDM					0.320	2.383	0.020
	Applicability					0.256	2.248	0.028
	Accessibility					-0.185	-2.113	0.038
	Model 2 (FDC added)	0.802	0.642	8.20%	0.000			
	Employee Development					-0.339	-2.583	0.012
	DDDM					0.257	2.066	0.043
	Coordination					0.317	2.136	0.036
	Accessibility					-0.170	-2.107	0.039
BDAC	Model 1	0.717	0.514	51.4%	0.000			
	Employee Development					-0.267	-2.081	0.041
	DDDM					0.369	3.079	0.003
	FDC					0.631	5.025	0.000
	Model 2 (BDAC added)	0.802	0.642	12.84%	0.000			
	Employee Development					-0.339	-2.583	0.012
	DDDM					0.257	2.066	0.043
	FDC					0.511	3.846	0.000
	Coordination					0.317	2.136	0.036
	Accessibility					-0.170	-2.107	0.039

Note: in each model, all factors other than the one added in model 2 were present in each case

Factors with significant correlation ($p < 0.05$) were illustrated per model. In the absence of OCBDA, FDC, strategic and accessibility were significant. When added in Model 2, strategic became insignificant, whilst FDC increase its weighting ($\beta = 0.371$ to $\beta = 0.511$), whilst

coordination also became significant and gained a large weighting at $\beta = 0.317$. It is interesting that 'Accessibility' returned such prominence in regression models when it was not significantly correlated to FPer and had a small correlation coefficient (Table 17; 0.051 and $p = 0.654$). This may be an artefact of its non-normal distribution (Table 16) and poor linearity (Section 9.4.8). Given this, and the low weighting in each result, its impact was disregarded from further application as it was unreliable.

Given the significance of positive impact of each of the assessed factors on the general model with respect to FPer, moderation of the BDAC – Fper capability can be confirmed, confirming H_5 , H_6 and H_7 through rejection of null-hypothesis in each case. Therefore the organisation does have a significant impact on BDAC, as expressed through FPer. Moderation of the BDAC – FPer relationship by each of the constructs was investigated through HMR, with potential for moderation illustrated through relative contribution to variance.

Chapter 6: Discussion of Findings

“The analysis of data will not by itself produce new ideas...” – Edward de Bono

6.1. Introduction

In this research, the primary aim was to gain a deeper understanding of whether the organisational environment, as described in this case by three constructs (OCBDA, FDC and DDDM) added to the base model in Chapter 3, influences the impact of BDA on FPer. Research Question 1 assessed the BDAC – FPer relationship as a confirmatory baseline. Research questions 2 and 3 assessed the individual relationships between FDC, OCBDA and DDDM on FPer and BDAC respectively, in order to characterise the impact of the organisational environment of the BDA environment and its outputs. In so doing this assessed the secondary intent of the research, which was to confirm whether the factors identified through research exhibited independent impact on BDAC and FPer respectively such that the basis for the primary aim could be confirmed. Finally, this study intended to characterise the relative impact of the organisational factor on the BDA environment, as addressed through Research Question 4, which also considered whether the organisational factors mediated the BDAC – FPer relationship. Furthermore, based on the above findings, this study intended to further characterise actions that should be taken by organisations to properly implement Big Data initiatives in organisations. This pertained to addressing how to make the organisation a more conducive environment for such a capability to be successful, and not just focussing on the technical attributes of a BDA environment itself, as is the case for a bulk of the literature on the topic (Chen *et al.*, 2014; Hashem *et al.*, 2015; Jin *et al.*, 2015; Liu *et al.*, 2016; Siddiqi *et al.*, 2016; Y. Wang *et al.*, 2015; Wang *et al.*, 2016; Yaqoob *et al.*, 2016; Zhou *et al.*, 2017).

6.2. Assessment of the relationship of constructs to FPer

5.7.1. Discussion of Research Question 1

Research Question 1: Is there a positive relationship between Big Data Analytics Capability (BDAC) and Firm Performance (FPer)?

Research question 1 sought to identify what respondents perceived to be the relationship between BDAC and FPer, such that the relationship between BDA as the core organisational capability being assessed could be directly associated with its intended result, which would be an increase in FPer for the organisation. This was fundamental to the study as it established the causal link between the capability and output. Furthermore, it illustrated that the BDA capability results in delivery of value to the organisation, and thus that is an area worthy both of being researched, and of being considered by business in order to support the capability of an organisation.

In congruence with Akter *et al.* (2016) and Wamba *et al.* (2017), BDAC was found to be strongly correlated to FPer (0.632 and $p = 0.00$). This thus confirms the base capability and assumption that BDAC does have a significant and a positive impact of FPer.

5.7.2. Discussion of Research Question 2

Research Question 2: Is there a positive relationship between Firm Dynamic Capability (FDC), which includes Firm Adsorptive Capability (FAC), Data Driven Decision Making (DDDM) and Organisational Culture to Big Data Analytics (OCBDA) respectively and FPer?

Though the organisational capabilities of FDC, DDDM and OCBDA have been studied extensively, the present study considers their direct impact on FPer through a model composition which has not previously used these factors. Akter *et al.* (2016) considered business strategy alignment in relation to the model, whilst Wamba *et al.* (2017) considered operational dynamic capabilities, a related but dis-similar composition to FDC considered herein. Furthermore, though the base interaction of interest in this study is the BDAC – FPer relationship, the organisational constructs added to the research model were each present within the organisations before BDAC was added. Accordingly, their relative individual relationship must be considered for relative comparison of their impact as these describe the impact of the environment in which BDAC is situated on FPer, independent of whether BDAC is present.

As presented in Table 20, strong correlations were found for each of the 1st order constructs with FPer, which, as the constructs were confirmed through CFA means the 2nd order constructs also correlated. FDC, planning and coordination presented with a strongest correlations for FPer. This is interesting as each of these are closely related with regard to impact on business as they all suggest the need to ensure that the organisation is flexible, the

changes are envisaged and acted upon by business in an orderly manner and that the business is able to enact such changes smoothly. In essence the planning and coordination could be assumed to be supporting capabilities of FDC as they enable the organisation to shift its capabilities such that it remains relevant in market (Côté-Real *et al.*, 2016; Mayer & Teece, 2008; Schilke, 2014; Teece & Leih, 2016). In close accordance with Teece *et al.* (1997) view on dynamic capabilities as being “*idiosyncratic factors which give rise to an organisations sustainable competitive advantage*”, the outputs of such a capability should be viewed as firm performance, thus it is not surprising that these correlate so strongly. This correlation is further validated by the variance described by loading this construct alone in model 1 of HMR, as this describes the independent level of variance that it accounts for with relation to FPer, shown in Table 20 as 44.5%. This is a significant portion of variance, and describes the nature of the theoretical description of FDC well, as it is a broad and encompassing theory that describes the dynamic of organisations reacting to shift in market based on their own capabilities.

Table 20: Summary correlations and contribution to variance by 1st order and 2nd order constructs

1 st order Constructs	Spearman's Rho coefficient		2 nd order Constructs	HMR		
	BDAC	FPer		Construct loaded in Model 1 R^2	Construct loaded in Model 2 R^2	ΔR^2
Support Future Capability	.528**	.391**	OCBDA	17%	59.9%	4.43%
Employee Dev.	.576**	.397**				
DDDM	.680**	.524**	DDDM	33.9%	62%	2.31%
FAC	.344**	.318**	FDC	44.5%	56.1%	8.2%
FDC	.661**	.636**				
Strategic Coordination	.607**	.436**	BDAMC	44.5%	50.3%	8.3%
Planning	.800**	.621**				
Implementation	.844**	.648**				
Applicability	.787**	.516**	BDAIF	28.9%	61.3%	2.9%
Accessibility	.334**	0.051				
Tech applicable to business	.656**	.341**	BDAPE	7.4%	57.3	1.0%
BDA Knowledge Transfer	.659**	.305**				
Firm Performance	.610**	-	FPer	-	-	-
BDAC	-	.610**	BDAC	52.2%	51.4%	12.8%

** . Correlation is significant at the 0.01 level (2-tailed).

Surprisingly, although correlations were not trivial (0.391 and 0.397), OCBDA received a comparatively low result. This suggests that organisations may not be highly conducive to adopting new innovations and capabilities, or that organisations do not quickly jump to adoption of new capabilities. Though this poses a barrier to adoption, it is not an entirely negative position as it suggests that organisations will adopt innovations, but only do so slowly, but would likely stay with them once they do. This is important as a capability such as BDA is not a trivial one, and also has broad impact across an organisation and would thus require integration throughout an organisation. Due to this, an effective adoption would require application of time, effort and patience by organisations to upskill to be able to engage with BDA as a tool.

DDDM resulted in surprising large correlations to FPer (0.524) given that on average, businesses do not have a well prepared access to data from which data driven business can prevail. Given that the respondents were, by default, part of organisation which had a strong analytics and data management system (as this is a prerequisite for BDA), the businesses may be predisposed to being data led. Data-based or evidence based decision making is significant toward extracting business value as it negates biases, such as those from decisions based on intuition, and also allows for decisions to become more consistent, cost effective, replicable, and can be transferred within the business (Brynjolfsson *et al.*, 2011; Davenport & Patil, 2011). Toward further application of BDA capabilities, this is promising as it describes a basis of organisations which are accustomed to replying on data for their decision making, which is how an analytics capability is able to deliver an impact on FPer.

In order for these capabilities to yield significance toward FPer, it is implicit that they must all be effectively deployed within an organisation. These are parameters which, if well implemented, allow for the innovations to take hold within business by overcoming implementation complexity (Comuzzi & Patel, 2016). Furthermore, they require that there is adequate leadership, goal setting and backing and management of initiative and personnel for decisions to be executed (Colas *et al.*, 2014; Lavallo *et al.*, 2011; Von Den Driesch *et al.*, 2015). Though the dimensions of OCBDA were relatively low (0.391 and 0.397), the ability to deploy innovations requires that organisational culture is conducive to them. As the underlying capabilities of organisations suggest that DDDM is already the implementation of a capability

such as BDA would mean that organisational culture is aligned with the embedded traits of the system, which further supports adoption (Leidner & Kayworth, 2006). Implementation of such capabilities also suggests that the organisations have structures and processes in place which are often reported at inhibitors of adoption (Chen *et al.*, 2014; Colas *et al.*, 2014; Lavallo *et al.*, 2011; Teece, 2010). Furthermore, it is suggested that businesses are expressing strategic alignment to implementation of BDA through these base capabilities being present and aligning to strategic positioning (Roelens *et al.*, 2017), or that they are sufficiently strategically flexible to adapt to dynamic changes in market (Schneider & Spieth, 2014).

The presence of these capabilities within organisation bodes well for ongoing adoption of BDA as organisations large enough to have BDA capabilities typically need to have the other base capabilities in place to operate on a normal basis.

6.3. Assessment of the relationship of constructs to BDAC

5.8.1. Discussion of Research Question 3

Research Question 3: Is there a positive relationship between FDC, DDDM and OCBDA and BDAC?

The link between an organisations environment and the functioning of capabilities such as BDA are well documented (Akter *et al.*, 2016; Alharthi *et al.*, 2017; Frizzo-Barker *et al.*, 2016; Wamba *et al.*, 2015). This study required that this be empirically asserted and described in relation to the BDA model applied herein. The direct correlations of FDC, DDDM and OCBDA respectively were assessed as separate as organisational characteristics in relation to on BDAC. This was done in order to describe the influence of the environment in which BDAC is situated on the ability of that BDA function to operate effectively. Given that BDA correlates positively with FPer, independent of other factors such as the business environment, a positive impact from the environment on BDAC is expected to further improve the impact of BDAC on FPer.

As presented in Table 20, strong correlations were found for each of the 1st order constructs with BDAC. Importantly, they were larger than the correlations to FPer in each case, which suggests that the business environment expresses a more direct impact on a functional capability like BDAC than it does on an output or result of business such as measured by FPer.

Similarly to FPer, the factors of OCBDA, DDDM and FDC all presented significant correlations, however they were all larger than 0.52 other than FAC which reported a correlation coefficient of 0.344 (BDAMC, BDAIF and BDAPE were not considered in this assessment as they form part of BDA, not the organisations environment). This illustrates the strong influence of the organisational environment on BDAC, which is significant within the context of this research as it affirms the need to study the influence of the business environment on BDA more thoroughly, as asserted by Comuzzi and Patel (2016). Furthermore, this illustrates the significant need for organisations to be made aware of the influence of key organisational attributes in order to make the business environment conducive to the adoption of a capability such as BDA. This is necessary as it does not matter how good a BDA capability is from a technical and functional perspective if the business, who need to action on analytics outputs from BDA, are not sufficiently engaged with the BDA capability and do not apply it in their respective roles.

The low correlation coefficient of 0.344 reported for FAC may be an attribute of FAC effectively being a narrower component of FDC, which was the strongest of the organisational influences at 0.680. Zahra and George (2002) proposed that absorptive capacity should be reconceptualised as a dynamic capability as that pertains to a firm's ability to both create and use knowledge to sustain competitive advantage, rather than to only incorporate outside sources of knowledge. Giniuniene and Jurksiene (2015) furthermore describe how dynamic capabilities (FDC), innovation (OCBDA) and organizational learning (FAC) are interrelated. This is significant herein as each are features which were assessed through this study, and all of which have been found to have significant impacts on BDAC. Furthermore, each of these components is regularly reported as having a significant impact on resulting firm performance. FDC thus provides a relatively broadly encompassing perspective of an organisation's ability to identify appropriate knowledge, initiatives, capabilities or processes, and then to adopt and implement them. In doing so it describes a higher level or more strategic link between an organisation's operations and its environment (Barreto, 2010).

As FAC as an independent factor relates to a firm's ability to identify existing external knowledge that is useful to it, and to then incorporate and apply this within the firm (Cohen & Levinthal, 1990; Elbashir *et al.*, 2011; Gao *et al.*, 2017; Greenhalgh *et al.*, 2004; Roberts *et al.*, 2012; Wisdom *et al.*, 2014), it describes an external view for an organisation to acquire

capability. Given this, it is central to initiating the process of allowing a capability like BDA to develop. It does this by strengthening or refocussing an organisations knowledge base and by increasing their ability to anticipate and appropriately position themselves to trends in market (Elbashir *et al.*, 2011; Roberts *et al.*, 2012). Given that FAC can be considered a component of FDC, their combined influence allows for the adoption and assimilation of technically complex capabilities, such as Big Data capabilities.

As described by HMR with each factor initially loaded by itself in Model 1 and then adding the rest of the environment in Model 2, their respective isolated variance can be compared. Therein, FDC explains a large portion of variance relative to the other at 44.5%. This is unsurprising within context of its influence on BDAC as it describes the ability of an organisation to not only identify capabilities, but to incorporate them into the operation such that it can maintain competitive advantage that can be expresses through FPer (Teece *et al.*, 1997).

The process of adoption or assimilation of capabilities such as BDA into organisations within an organisation is well documented (Elbashir *et al.*, 2011; Rad *et al.*, 2017; Wisdom *et al.*, 2014). Though FDC provides a description of an organisation level functional capability; OCBDA describes the perception and norms of the people in an organisation, as well as the organisational culture towards a capability which impact on the ability for FDC to be executed. Appropriate organisational culture toward innovations is widely attributed as being a pillar of organisational success (Donate & Guadamillas, 2010; Guerra & Knox, 2008; Leidner & Kayworth, 2006; Martín-de Castro *et al.*, 2013; Rogers & Meehan, 2007; Shahzad *et al.*, 2017). With correlations of 0.528 and 0.576, OCBDA was found to be strongly associated with BDAC. Importantly, it was more strongly correlated to OCBDA than it was to FPer (0.391 and 0.379). This is unsurprising as OCBDA effectively describes the openness of an organisation to engage with innovations. This would thus be expected to have a higher correlation with a capability such as BDA than with FPer, which is as a result of BDA. Furthermore, Elbashir *et al.* (2011) describe that operational users of capabilities such as BDA within business directly impact the manner in which systems are deployed, thus the impact of OCBDA would be expected to have a strong correlation with BDA, which has direct operational outputs, with many examples of operational efficiencies having been reported from BDA (Chen *et al.*, 2012; Lavalle *et al.*, 2011; Lee, 2017; Mazzei & Noble, 2017; Pospiech & Felden, 2012). OCBDA accounted for only 17% of

variance when loaded in Model 1 of HMR, indicating that, similarly to the situation reported for FPer, though it is a significant requirement, it is not strongly influential in this model beyond being a required base capability. In the global business environment of today, which is characterised by rapid development of technological capabilities and an ever increasing dependence on technologies, it is surprising that organisations are not more conditioned to being open to new, and particularly technology related innovations, especially when it is taken for granted that being innovative and making use of technology is crucial to retain market competitiveness (Ben-Menahem *et al.*, 2013; Roberts *et al.*, 2012). This study is grounded in the assumption that the global business environment is becoming ever more complex, ever more reliant on technology, ever more challenging to navigate, and ever more in need of solutions to assist organisations deal with the wave of requirements coming at them. As a result of this, the rapid increase in the availability of data and the expectation for quick reaction by market, the ability to be efficient and effective in assimilation of innovations, knowledge and capabilities to maintain relevance in market is becoming ever more essential (Barreto, 2010). The Big Data environment is the interface of multiple rapidly evolving and cutting edge fields which, given its attributes, has the potential to assist with many of these issues. The strong FDC correlation, along with FAC and OCBDA is promising in these results as it describes an organisational prerogative to assimilate a capability that itself if not trivial.

Importantly for this perspective, DDDM had the largest correlation coefficient (0.680) of all of the constructs assessed, which illustrates the critical path that DDDM has in facilitating BDAC. This is an entirely logical result as DDDM describes businesses propensity to make decision based on data rather than on gut feel. As such a higher propensity for DDDM translates to a higher application of BDA's outputs, which is the only way that BDAC is able to impact on FPer. This strong result was further validated by DDDM accounting for 33.9% of variance in Model 1 of HMR. What is of further significance from an environmental perspective with the prominence of DDM is that the organisational environment impacts on establishment, inputs, and outputs of the BDAC, and this BDAC is reliant on the organisational environment facilitating it as much as it intends to facilitate the organisation to deliver improved FPer.

6.4. Assessment of the relative weighting and impact of constructs on the BDA – FPer relationship

6.4.1. Discussion of Research Question 4

Research question 4: What are the combined impacts of FDC, DDDM and OCBDA (independent variables) on the BDAC – FPer (dependant variable) relationship, and do these independent variables moderate that relationship?

Given the nature of these constructs, FDC, DDDM and OCBDA are capabilities within an organisation which are anticipated to impact on FPer through their direct impact on the BDAC function, and through this moderate FPer. Investigation of the extent of impact through variance explained in a model, as well as the direction of influence allows for the interactions to be characterised, and the effect of the organisation on the BDAC – FPer relationship to be characterised more thoroughly. In so doing, this aims to describe whether the environment of the organisation facilitates BDA as a capability, and its ability to deliver value to the organisation can be enabled through influence of the organisational attributes (environment). This study assessed this relationship by investigating the combined impact of FDC, DDDM, OCBDA and BDAC as organisational capabilities (independent variables) and FPer (dependant variable) in order to assess their relative contribution to variance within the model.

As summarised in Table 20, Organisational constructs were added in Model 2 in HMR, independently of those in model 1, allowing for their relative impact to be illustrated through the change in variance which the result in, ΔR^2 . This thereby described the influence of each of the organisational constructs relative to one another, and thereby illustrated their relative impact through variance R^2 , which was found to be 64.3% for the entire model, a fair result for a model that considers such complexity in the factors assessed.

6.4.1.1. Assessment of organisational constructs on BDAC – FPer

BDAC was assessed to describe its independent contribution, and relate it in context back to research question 1. Through HMR, BDAC was found to resulted in the highest $\Delta R^2 = 12.8\%$. This illustrates that it contributes most significantly to FPer in the model by accounting for the highest proportion of variance. This describes the direct relationship between BDAC and FPer, and supported the position that the organisational constructs support the functioning of BDA, which in turn delivers FPer, and thus that this relationship is moderated by the organisational

capabilities. The results therefore support the findings across the model, with the primary relationship being BDAC – FPer, but with organisational parameters impacting on both of these parameters. This in turn describes that this primary relationship is moderated by the organisational constructs. The individual impact of each of the organisational variables on the total variance in the model was illustrated through their respective ΔR^2 , wherein OCBDA $\Delta R^2 = 0.443$, FDC $\Delta R^2 = 0.231$ and DDDM $\Delta R^2 = 0.0.820$, as summarised in Table 20.

Of interest from these results is that the higher weighting of variance is explained by the constructs which describe inputs to the BDA environment (OCBDA and FDC) rather than in the form of outputs from it (DDDM). This is not an unsurprising result given the early stages of development of the BDA capabilities within the sample set, wherein 36% of respondents (Table 11) and 52% of organisations (Table 8) were within their first two years of exposure to BDA. This suggests that the primary influence would be the effective assimilation of the BDA capability into the organisation. This may thus be a factor of BDAC capabilities having not yet reached a stage of maturity where their output (insights and recommendations from data) are of significance. Importantly however, though the base values are fairly low, they do describe the definite impact of the organisation on firstly BDA, and then on the BDA – FPer relationship. This therefore confirms that the organisation in which a BDA capability is implanted will have a direct impact on the ability of the function to assimilate itself. Secondly, as the application of its outputs are through business, it is confirmed that business must also facilitate the.

6.4.1.2. Understanding the BDA capability construct

The composition of the BDAC construct within the model was further validated by assessing the 2nd order constructs of BDAMC, BDAIF and BDAPE respectively. The BDAC construct has previously been shown to have a positive association with all of the level one constructs or primary dimensions that are described in Figure 1 (Akter *et al.*, 2016; Wamba *et al.*, 2017). Though terminology differed slightly between the two, the level two constructs included BDA Talent Capability or Personnel Expertise, BDA Technical Capability or Infrastructure Flexibility and BDA Management Capability. In each case these constructs were found to be the three primary components of a BDA capability.

Akter *et al.* (2016) reported that Talent Capability was the highest weighted factor (BDA Talent Capability $\beta = 0.949$, $R^2 = 0.900$; BDA Management Capability $\beta = 0.938$, $R^2 = 0.879$; BDA Technical Capability $\beta = 0.924$, $R^2 = 0.854$). Wamba *et al.* (2017), similarly found that Personnel

Expertise (Talent Capability) was the highest weighted factor (Personnel Expertise $R^2 = 0.93$, Infrastructure Flexibility $R^2 = 0.92$, Management Capability $R^2 = 0.86$). In both cases this suggests that the best gains in overall BDAC are achieved by having the most capable employees within as skill sets within the environment are a premium resource. In this study, CFA confirmed the appropriateness of the third order factor, BDAC as a cumulative result of BDAMC, BDAIF and BDAPE (Section 5.4, Figure 13) thereby allowing for it to be assessed directly as a construct. Similarly to both Akter *et al.* (2016) and Wamba *et al.* (2017), this study found that BDAIF and BDAPE were closely weighted phenomena (BDAIF $R^2 = 0.61$; BDAPE $R^2 = 0.57$; BDAMC $R^2 = 0.50$). Variation was not significant as reference studies reported their interaction model explaining around 60% of the variance, whilst 64.3% was accounted for in this study. accounted for was lower than that described by either Wamba *et al.* (2017) or Akter *et al.* (2016), and is expected to be as a result of several factors. Firstly, this study considered several exogenous impacts on the model which themselves are influences by multiple parameters with the potential for multilinearity and interdependances on the constructs describes, as well as others. As such, the models presented by Wamba *et al.* (2017) and Akter *et al.* (2016) were required to account t for lees variation than was strived for in this study. Secondly, the sample applied in this study was significantly smaller than the previous two studies, and thus the resulting impact may have been modelled more accurately with a larger sample set that would have minimised bias. Thirdly, SEM was applied in the reference models, which allows for a more complex relationship to be accounted for than by HMR, as was applied in this study.

The findings in this study do however concur with the findings of Akter *et al.* (2016) and Wamba *et al.* (2017), as well as the broadly reported position form literature that within the BDA capability, the technical capabilities of the BDA platform, as well as the personnel who have very rare skills sets at this time due to the newness of the capability are among the most influential factors (Khan *et al.*, 2017; Larson & Chang, 2016; Miloslavskaya & Tolstoy, 2016; Shin, 2016; Sivarajah, Kamal, Irani, & Weerakkody, 2017; Yaqoob *et al.*, 2016). This is not surprising given the highly mathematical and technical nature of the Big Data and BDA environment, as this is essentially a skill set that requires a combination of statistics, computer coding and IT systems engineers, each of which by themselves are complex skill sets. This too is in agreement with Pospiech and Felden (2012), who assessed Big Data literature and found that 87% is technical in nature as a direct result of the characteristics of the environment. As a

core technical function of a big data capability is the assimilation of vastly different forms of data from multiple sources, in high volumes, and often at high speed, infrastructure must be able to enable this of the BDA capability will not be able to function (Larson & Chang, 2016; Miloslavskaya & Tolstoy, 2016; Shin, 2016; Sivarajah *et al.*, 2017; Yaqoob *et al.*, 2016).

Furthermore, the BDA capability is still in its infancy, having only gained prominence since around 2011 (Lee, 2017). This also corresponds well with the descriptive analyses of the sample set, where 36% (Table 11) were within their first two years of exposure to BDA, and 52% (Table 8), of organisations were within their first two years of exposure to BDA. This suggest that the capabilities are still very new within organisations, and the predominant view of those respondents and the organisations will be the technical setup of the platform, as well as the ongoing struggle to find appropriately skilled people, particularly data scientists, data engineers and IT systems engineers who are skilled in Linux and management of virtual deployment of machine infrastructure.

It is interesting that unlike in Akter *et al.* (2016) or Wamba *et al.* (2017), where BDAMC was found to have the highest correlation coefficient (BDAPE = 0.722, BDAIF = 0.799, BDAMC = 0.907). This may also be a factor of the functions within many of the organisations assessed having not yet reached sufficient maturity for the impact of BDAIF and BDAPE to be perceived to have sufficient weighting, wherein if in the very early stages, the impact of BDAC would be limited to the team developing it. This is a feasible explanation given that only 30% (Table 10) of respondents were directly associated with analytical or technical attributes of the BDA capability respectively. Given this, respondents may have perceived the capability from a business perspective, where if they were not intimately familiar with development of early stages of the capability, they may have then up-scaled BDAMC. This may also have been an artefact of the total question set in the survey, wherein organisational attributes (OCBDA, DDDM, FDC) were questioned before to BDAC in the sequence of the survey. In the reference studies, the opposite was true, which may have shifted respondent's perceptions to view the capability from an external or rest-of-business perspective, and not from an internal perspective of the BDA capability as may have been the case in the previous studies. This is bolstered by BDAMC being the highest weighted factor, which supports the proposition that the BDA capability may have been viewed from an organisational perspective as the construct

relates to management, planning and implementation of a capability more than it does to technical attributes of data science itself, where BDAPE would be more relevant.

The weighting of BDAMC in this study may indicate that achieving sustainable deployment of a BDA capability and translating that into competitive advantage for the organisation relies heavily on decision makers. This may be particularly true in the early stages of the BDAC functions establishment, as a significant input of effort, time and alignment of diverse entities within the business is necessary, which is a feature which cannot be underestimated, and is something that requires good management and strong leadership. Several studies, which include extensive market surveys and focussed theoretical assessments, report the key importance of having quality managers in elevating or enabling BDA capabilities within the organisation, particularly where the establishment of the function within a business is involved (Brown *et al.*, 2011; Espinosa & Armour, 2016; Gupta & George, 2016; Kane, Palmer, Phillips, & Kiron, 2015; Kane, Palmer, Phillips, Kiron, & Buckley, 2016; Kiron *et al.*, 2014; Lau *et al.*, 2014; Lavallo *et al.*, 2011). These describe the very human attributes of business capabilities which must often be actively and delicately planned for and managed, particularly if the implementation disrupts the business and requires change management or influences business strategy on focus (Gupta & George, 2016; Mazzei & Noble, 2017).

As all were large, significant and not entirely different to those reported by Akter *et al.* (2016) and Wamba *et al.* (2017), the presence of each was considered to be of more importance than the relative weighting. Furthermore, as each function is of key importance to BDA function as a whole, there appears to be limited benefit in assessing which is more important, as they must all operate effectively for BDA to be possible. This is further emphasised through assessing the relative impact of organisational impacts through the remainder of this section. Thus it is recommended that business pay equal attention to all the dimensions of the environment to ensure effective application.

6.5. Assessment of BDA in the business context

6.5.1. Apprehension to BDA by people in the business environment as a result of a lack of understanding

Pockets of extremely positive results have been reported in business which, along with the significant potential of BDA capabilities, has fuelled the drive to develop the BDA capability

further (Chen *et al.*, 2014; Hashem *et al.*, 2015; Jin *et al.*, 2015; Liu *et al.*, 2016; Siddiqa *et al.*, 2016; Wang *et al.*, 2015; Wang *et al.*, 2016; Yaqoob *et al.*, 2016; Zhou *et al.*, 2017). This is a factor of the broad perception in market that BDA is a capability that business must capitalise on. The corollary of this very positive disposition on the BDA environment is that the fact that there are pockets of capability means that BDA as a capability is typically not homogenously distributed through businesses. Furthermore, although businesses ‘want’ BDA, business people are typically insecure about it, concern about what it will do to their businesses, how it will impact on employees and where their markets are going to shift as a result (Frizzo-Barker *et al.*, 2016; Jin *et al.*, 2015; Kaisler, Armour, Espinosa, & Money, 2013; Özköse *et al.*, 2015). Technical difficulties too are well documented, with a majority of literature tackling these issues (Jin *et al.*, 2015; Kaisler *et al.*, 2013; Özköse *et al.*, 2015; Singh & Reddy, 2014).

It is suggested that the source of this trepidation by people in business is due, in large part, to two main sources which contribute to a lack of understanding of BDA. The first is that the capability is still in its infancy, and thus has not yet been well described and characterised to make is more intuitive to people not intimately involved with the environment (Özköse *et al.*, 2015; Wang *et al.*, 2016).

The second may be derived from the fact that BDA capabilities have primarily been advanced through businesses, and as such there is a degree of disparity in terminology, supported by a lack of clarity that is a result of organisations not disclosing all of their IP, as this too is a source of business potential in its own right, and of competitive advantage (Mazzei & Noble, 2017; Özköse *et al.*, 2015; Shin, 2016). Examples of these businesses include Google, eBay, Amazon, Walt Disney, Netflix, United Parcel Services (UPS), Airbnb, Spotify, Yahoo! and Facebook to name but a few (Alharthi *et al.*, 2017; Hashem *et al.*, 2015; Yaqoob *et al.*, 2016). Closely related to this is that it is typically viewed as a ‘black box’ which can only be appreciated by the Data Scientists, who it seems, may be considered in a light where is assumed they are remarkably intelligent, and that what they do will as a result be unintelligible to an ‘average’ person in business. This generally has resulted in a focus on technical details of the BDA environment, which are not implicitly obvious to people who do not deal with it regularly, rather than how to transpose BDA as a capability into an environment which did not develop the capability itself, and thus one which may not be fully prepared to assimilate it. The implication of which is that business people are not prepared for the capability and are hindered from actively

embracing it as they do not understand it, which typically limits acceptance (Leidner & Kayworth, 2006; Loebbecke & Picot, 2015; Shah *et al.*, 2017).

This is significant as this study illustrates the clear and significant influence of the organisation, and thus the people that work in the organisation, on BDA. It is the strong position of the author that this is a significant disservice both to BDA as a capability and to business in general, as it distances people in business from trying to understand BDA. The result of this is that it limits the ability to effectively apply BDA, both because business people will not know what to ask of the BDA capability, and because they will not know how to apply outputs of that capability. If business don not understand BDA then they will not grasp it properly. If they do not grasp it, then they will not apply it. If they do not apply it, then the BDA capability will continue to deliver limited FPer as it does not matter how good an insight generating capability is if it cannot be executed by that business. In order to combat this, and to improve to inter-relation of BDAC with the organisation, and hence to deliver improved FPer, it is essential to make BDA a more intuitive for people in business for it to be embraced as a functional tool, rather than a distant potential. This need is further emphasised by the fact that in the face of prevailing dominant logic, people conform to norms and beliefs to guide their action, which will be further emphasised if it is difficult to understand an innovation, as people will not give it a chance (Wrigley *et al.*, 2016).

6.5.2. Engagement by BDA with business to improve output

Application of a data driven decision making process has been shown to directly improve daily business function through better, more timeous and more stable decisions being taken (Brynjolfsson *et al.*, 2011; Lavallo *et al.*, 2011; McAfee & Brynjolfsson, 2012; Nannetti, 2012). Though being good at this can be a competitive advantage, this is a base capability which must be present in any successful business. This in turn means that data and analytics should be strategic capabilities, central to business operations, which it often is, at least to some extent. Although BDA allows for immense data sets to be analysed, and that it does so with technically complex systems and methodologies, the output of BDA in the form of insights, as well as the inputs to BDA in the form of business problems should not be viewed by business as anything different to typical analytics. This is because BDA as a process still requires a target against which it builds its analytical models and processes the data (Colas *et al.*, 2014; McAfee *et al.*, 2016; Provost & Fawcett, 2013). The complexity of the BDA environment should not be

allowed to translate to the business environment as the operational components of business. This is because the operational components of business are typically the sources of the business problems that need resolution, and are also the effective mechanism through which outputs are enacted. The direct implication of this is that BDA must position itself as a business enabler which resolves operational capabilities early in its lifecycle, so that people in the business environment understand it and its application, as it will then be in line with their prevailing dominant logic (Wrigley *et al.*, 2016), and be positioned as a level where it is intuitive. Importantly, this also allows for so called ‘quick wins’ as these issues are often not computationally complex, but allow for the setup of data feeds, infrastructure and processes, which are the BDAIF and BDAPE components of the BDA model, which were found to be most significantly impactful in this study, as well as those conducted by Akter *et al.* (2016) and Wamba *et al.* (2017).

Once these capabilities are enabled, the BDA can expand the computational and influential aspects on business, both through having developed its own capabilities, and through having established working relationships in the organisation. Taking this approach is significant as Comuzzi and Patel (2016) reported that up to half of all Big Data projects in large organisations are never completed.

Chapter 7: Conclusion

7.1. Introduction

In this Chapter, the influence of the organisation on BDAC is discussed with relation to the research model applied in in this study. The implications of the direct analyses will then be discussed to distil the key contributions and recommendations that can be derived from this body of work for theoretical and practical applications in business in business situations respectively. Recommendations toward future research are also presented in relation to the body of work done.

7.2. Contributions made toward theory by the research study

The study built on the foundation of research which has begun to investigate the BDA environment within organisations, and in particular its influence of FPer (Akter *et al.*, 2016; Kim *et al.*, 2012; Wamba *et al.*, 2017). A core feature of these analyses was to move toward answering what capabilities and characteristics an organisation should portray or develop, whether technical or not, in order for BDA capabilities to be successful. The study moved to provide sound empirical findings to substantiate this question, and furthered the existing body of knowledge by considering further dimensions of organisational capabilities, and the impact that they have on deployment of BDA as a capability.

From the perspective of the total environment assessed, the study presented a model which illustrated the interactions between 14 constructs in order to explain the variance observed in FPer as a result of the interactions of organisational attributes and BDAC. This was addressed in relation to the influence of a broader spectrum of organisational constructs than has been tested previously, and included aspects that pertain to implementation of BDA in an organisation (FDC and OCBDA) and well as application of BDA through DDDM. 64.3% of variance was accounted for in this study, which was in congruence with previous models (Akter *et al.*, 2016; Wamba *et al.*, 2017) which also reported in the region of 60% of variance explained. Unlike Akter *et al.* (2016) and Wamba *et al.* (2017), 9 rather than 11 1st order constructs were found, however equivalent 2nd order construct loading to that presented in the previous studies was seen. This is significant as it further verifies the applicability of the model itself, as in this case it was applied in a predominantly South African context, whereas

the previous two were applied in the USA and China respectively. Importantly, data collected for empirical analyses were collected from across 12 different groups of industries, and well as broad experience of respondents in BDA. Furthermore multiple perspectives from within the business environment were garnered, which allows for a balanced view of the capability to be derived. Furthermore, these findings agreed well with Kiron *et al.* (2014), who similarly to Akter *et al.* (2016) and Wamba *et al.* (2017) report that the core capabilities within a BDA function itself are effective capabilities in technology, management processes and significantly for early stage deployments, talent.

Although the importance of these three BDA capabilities are relatively well established (Akter *et al.*, 2016; Chen *et al.*, 2012; Kiron *et al.*, 2014; Wamba *et al.*, 2017), this study was among the first to assess the impact of BDA on FPer, and, to the best knowledge of the author, is the first one to directly consider the parameters of OCBDA, FDC and DDDM in relation to the BDAC – FPer relationship. This thus highlighted the role of FDC, OCBDA and DDDM in the process to further describe the role of the organisation on the functioning of the BDA capability. The complexity of interactions was illustrated through direct correlations between each being found in relation to BDAC as well as FPer, as well as moderation of the direct BDA – FPer relationship. This lends further backing to the assertion that the environment being assessed is highly complex and interconnected, however also describes the need for further research in the area to appropriately characterise interactions.

This study builds on the theory of BDA strategy, and provides insight toward how to leverage the BDA as a function within the organisation in order to support improved FPer. It does this through extending the notion that the organisation (FDC, OCBDA, DDDM) is a central component (in addition to BDAPE, BDAIF and BDAMC) of the ability of BDAC to deliver value to the organisation through FPer. Importantly, these organisational attributes characterised aspects of interaction from acceptance of an innovation (OCBDA) to implementation of an innovation (FDC) and through to application of an innovation (DDDM), each of which was found to have significant implications. The relative influence of these capabilities was also considered relative to FPer and BDAC in their individual capabilities and through moderation of the BDC – FPer relationship. Overall, these findings suggest that organisations which intend to apply BDA should consider organisational alignment to BDA as a crucial antecedent to establishing a BDA capability if influence on FPer is to be achieved.

7.3. Practical implications that can be relayed to business

With the backdrop of a sustained growth in interest in BDA capabilities by business, due to its potential to support multiple capabilities across business and to support sustained competitive advantage, this study advances BDA as a practically applicable medium to support improved FPer. These findings are of relevance to practitioners such as BDA managers, senior executives, and to consultants who may be engaged in the implementation of BDA capabilities within organisations. The key attributes which are relevant toward this are that the broader business environment do impose a significant impact on the ability of BDA capabilities to function within the business, and thus that the BDA capability cannot be treated as a standalone function, but rather that the organisation should be strategically aligned to BDA if it is to deliver on its promise.

The presence of moderating impact of OCBDA, DDDM and FDC on the BDAC – FPer relationship provides business with an avenue through which sustainable competitive advantage may be achieved. This however requires that the organisation be aligned to the deployment of BDA. Importantly, it is important to note that significant correlations were found for factors which relate to acceptance of an innovation (OCBDA), implementation of an innovation (FDC) and through to application of an innovation (DDDM), this alignment must be put in place for the BDA capability to be deployed effectively, but must also be maintained for it to deliver sustained outputs in the organisation. If the organisation was to not be aligned, or these capabilities were not in place, BDA as a capability may not gain, or may lose its effectiveness. Given the current pressures and dynamism of market in general, BDAC provides a potential solution, however the above describes that the organisation must commit itself to BDA in order for it to draw value, otherwise it has the potential to be an expensive activity with limited FPer returns. Doing so will however require that managers concentrate on all three sub components of BDA, (BDAC, BDAIF and BDAPE) as well as to the interaction of the organisation with it (as described in this case by FDC, OCBDA and DDDM). Given this complexity, and the challenges business will face in executing this in such a complex environment, as was found in this study, the requirement for quality BDA managerial competencies cannot be underestimated, both to facilitate the BDA capability itself, and to enable alignment of BDAC with business.

Achieving this should include a significant effort to better relate the BDA capability to business, such that business can understand the capability, and not be intimidated by it. Importantly, the complexity of the technical concerns of a BDA environment should hold no bearing on the way the organisation views it. Ideally, the organisation should see BDA as just an iteration of an analytics capability with which they are familiar. The implication of which being that the BDA capability will sit within the norms of activity, and will thus be accepted and applied more readily by the organisation. This is significant as the organisation are not only the source of the business problems which BDA will endeavour to solve, but they are also the conduit through which the outputs of BDA are actioned. Thus, for BDA to be successful, its interactions with the rest of the organisation cannot be underestimated.

7.4. Recommendations for future research

Though this and the preceding studies have provided a strong basis of empirical evidence to describe the interactions of BDAC and characterise the BDAC model, the complexity of the environment being assessed cannot be underestimated, and thus there is significant potential for further research which assesses further dimensions of the impact of the organisational environment on the BDA capability. Suggestions toward this are listed below;

- The organisational constructs assessed are not exhaustive of the parameters which may influence BDA capability being effectively rolled out in an organisation. Further research should consider implications of further factors that are internal to the organisation, as well as external factors to the organisation which may impact on the capability. These may include, but not limited to; legal considerations, governance, data capability, talent and development, culture, appreciation of scientific method by business people, executive sponsorship, strategic alignment of business with big data platform, strategic objectives, knowledge development of the rest of the organisation with respect to BDA and DDDM, means of targeting what projects will deliver the greatest value for a given organisation relative to BDA, what application of BDA techniques best align to different business environments and what curation and presentment of data facilitates better uptake by business, no name but a few. Though each of these should be studies with regards to direct impact on BDA, the impact in FPer, either directly, by moderation or by mediation of the BDAC – FPer relationship must all be described such that they can be leveraged.

- Though a small number of international responses were acquired, the sample frame was primarily localised to South Africa. This may present a bias toward perceptions of BDA and its influence on FPer as a factor of cultural norms within business in this country. As the impact of the business environment is of such significance on the ability of a technical capability to achieve effective deployment, the relative organisational cultures and norms may present significant insight in how to better implement these capabilities in organisation around the world. This is also relevant to large organisations which may have divisions that compartmentalise capabilities and perceptions. These may have had success in deployment in areas of the business, but have battled to deploy it across the business units or across localities of the business, as may be the case for multinational corporations.
- To address the concern of non-response bias, profiles of survey's respondents could be checked with regard to size of organisation, industry exposure to BDA and degree of maturity of deployments of BDA within the South Africa. Through this, non-response bias could be assessed through chi-square tests.
- The potential for lack of understanding of the BDA environment by persons in the business environment could also be addressed in order to determine how to more appropriately explain the capability so that it can be understood, and also be made more accessible as usable output by people in market.
- In CFA, the divergence of the first order item loadings, which included some low values suggests that the model applied may require further investigation in order to be more broadly applicable. Though Akter *et al.* (2016) and Wamba *et al.* (2017) respectively reported positive high average item loadings and a narrower range of difference, which was illustrative of respective items having greater convergence in measuring the underlying constructs presented, with minor changes in the construct composition in this study, this loading was significantly impacted. Though the sample size of this study is likely to be a significant factor in driving this result, a fragility in the model is suggested, which could be dispelled with further assessments concerning model composition, in order to develop guidelines of what must be included in future studies.
- Given that the factors assessed have inter-relation of impact, it would be prudent to assess whether there was a degree of common factor loading that may skew the total model output. This could be done through common method variance by applying Harman's one-factor test.

- Towards addressing concern on the broad ongoing applicability of the model, its predictive ability across more situations could be assessed (Stone–Geisser's Q^2) would further verify the appropriateness of the base model, and may possibly allow for re-assessment of its components and the potential for it to be more flexible, potentially also with a smaller number of questions asked.

7.5. Limitations of research

The limitations below describe the caveats and potential shortcomings identified for this study, owing to the choices made in research design, sampling, data curation and other potential biases that may have arisen through this study;

- The constructs assessed in this study are not exhaustive with respect to potential parameters which may influence BDA capability being effectively rolled out in firm or on the relevant impact of the relationship between BDA capability and FPer. Additional constructs with potential implications as determined from literature may include, but not be limited to; legal considerations and influences of ethical governance (such as considerations pertaining to acquisition, storage and application of certain data towards business ends), data capability, talent development and acquisition, development of culture that is actively inclusive of such capabilities, appreciation of scientific method by business people, executive sponsorship, strategic alignment of business with BDA.
- Though a small number of international responses were acquired, the sample frame was primarily localised to South Africa. This may present a bias toward perceptions of BDA and its influence on FPer as a factor of cultural norms within business in this country. These considerations cannot be considered herein due to the absence of appropriate sample data.
- South Africa is a relatively small market from a BDA perspective and thus;
 - There are limited organisations on which data would realistically have been acquired. Thus, given that the unit of analysis is of the organisation, the relative sample of organisations assessed would be relatively limited.
 - A majority of the organisations which do have sufficient data to establish BDA capabilities appear to be relatively immature with regard to implementation of BDA as a business capability. Thus the reported perceptions of both impact and implication of BDA on FPer may not be reported by respondents relative to known

mature BDA deployments in business, (i.e. respondents may not know what a good influence of BDA on FPer really looks like).

- Although respondents were selected through multiple methods in order to reach sufficient appropriate sample size, sampling biases may still be present;
 - Respondent's availability or will to complete the survey questionnaire may well have been a function of a combination of how busy they were, their desire to complete questionnaires given the number they are exposed to, being a relatively small sample from which to draw, and the depth and breadth of the questionnaire meaning that it was perceived to take a long time to complete. This is expected to impact primarily on CxO level personnel.
 - Given the highly mathematical and technical nature of the big data environment, there appears to be a perception in market that BDA may be conceptually out of reach of a majority of people, where in reality, unless one is a data scientist, there is no need for this complexity to be transferred to people outside of the BDA environment. Unfortunately, these are the very people who appear to be hindering the current application of big data platforms as they are not actively driving appropriate business models innovations to effectively incorporate these platforms. Furthermore, this is likely to have limited the response rate. This is likely to have arisen through limited appreciation of the big data environment by a vast majority of people leading to a perception of it being more complex than it should be. The implication of which is that people shy away from engaging, which is the opposite of what is required as described in this study. As a result, active strides need to be made to position BDA as a tool, wherein the complexity of the analyses themselves are divorced from the output.
 - Sample size both as a total and then further emphasised per strata (descriptive divisions or characteristics of respondents) was on limited which would limit the ability to assess, or extrapolate upon data associations observed. Though the impact of this was actively sought to be minimised through trying to gain as large a sample as possible, this remains a relevant consideration within this study. Extrapolation of the results should thus be done with due caution
- Theories being applied may require further development given limited empirical research with regard to BDA. This research only considers select constructs, and is not an exhaustive

assessment of the potential interplay between them and their influence on BDA, FPer or their relationship, thus further assessments may be required based on the outcome of this study.

- As the factors being assessed are complex, relatively under researched and their modes of inter relation with regard to the subject matter are not common knowledge within business environments, they may not have been fully considered by respondents, and may also not be obvious factors which may influence traits identified in analysis of interviews.
- The author of this study acknowledges fallibility in not being a seasoned researcher within this field, as previous experience was centred within the absolute measures of positivism in analytical scientific research. This in itself could have proven problematic in the research process as it pertained to establishment of constructs, their specific questions and combinations thereof in the questionnaire.
- As the chosen approach was a quantitative, positivist research design, and applied a self-administered questionnaire, the questionnaire provided a fixed list of answers from which a respondent was required to select. Though the constructs and questions applied in the questionnaire were derived from literature;
 - They may themselves not have been exhaustive assessments of the constructs given the need to curtail survey length to promote number of responses.
 - They may not have appropriately represented the required responses given that a new environment as assessed and the questions were derived from studies on alternate fields.
 - Options provides as answers may not have appropriately captured perceptions or sentiment of respondents in relation to each question.
 - Given the complexity of the environment, the reality of the BDA environment of an organisation may not have been appropriately characterised, as was reported on
- Biases in sampling may furthermore have resulted as a factor of non-probability and snowball sampling methods, wherein all appropriate employees in the organisation for which responses were gained may not have been engaged with a survey as they were outside of the sphere of influence of the researcher (Zikmund *et al.*, 2012).
- A potential bias may have arisen through answers of respondents themselves;

- Though the survey was distributed to a large number of people, the non-homogenous group of responses as a result of respondent non-response may result in a non-representative sample (Zikmund *et al.*, 2012).
- Bias may have occurred through inaccuracy or dishonesty of respondent answers to questions posed (potentially through factors such as guessed responses or a will to answer questions, even if they did not actually know the answers). This may thus result in a data set that may not be representative (Saunders & Lewis, 2012).

References

- Accenture Digital. (2015). Analytics maturity assessment. *Accenture Digital*, 2015. Retrieved from https://www.accenture.com/_acnmedia/Accenture/Conversion-Assets/DotCom/Documents/Global/PDF/Digital_3/Accenture-Analytics-Maturity-Assessment2015-the-Netherlands-Accenture-Digital.pdf
- Adcock, R., & Collier, D. (2001). Measurement Validity: A Shared Standard for Qualitative and Quantitative Research. *The American Political Science Review*, 95(3), 529–546.
- Addo-Tenkorang, R., & Helo, P. T. (2016). Big Data Applications in Operations/Supply-Chain Management: A Literature Review. *Computers & Industrial Engineering*, 101. <https://doi.org/http://dx.doi.org/10.1016/j.cie.2016.09.023>
- Agostini, L., Nosella, A., & Filippini, R. (2017). Users and radical innovation performance: the moderating role of the organisational context. *Technology Analysis & Strategic Management*, 28(7), 798–810. <https://doi.org/10.1080/09537325.2016.1156667>
- Agresti, A., & Franklin, C. (2007). *Statistics: The Art and Science of Learning from Data*. New Jersey: Pearson Education, Inc.
- Akter, S., Wamba, S. F., Gunasekaran, A., Dubey, R., & Childe, S. J. (2016). How to improve firm performance using big data analytics capability and business strategy alignment? *International Journal of Production Economics*, 182, 113–131. <https://doi.org/10.1016/j.ijpe.2016.08.018>
- Alharthi, A., Krotov, V., & Bowman, M. (2017). Addressing barriers to big data. *Business Horizons*, 60(3), 285–292. <https://doi.org/10.1016/j.bushor.2017.01.002>
- Baines, T. S. L. H. (2014). Sustainable business models. *Emerald Insight*, 29(5), 494–519. <https://doi.org/10.1108/JFM-03-2013-0017>
- Baldassarre, B., Calabretta, G., Bocken, N. M. P., & Jaskiewicz, T. (2017). Bridging Sustainable Business Model Innovation and User-driven Innovation: A Process for Sustainable Value Proposition Design. *Journal of Cleaner Production*, 147, 175–186. <https://doi.org/10.1016/j.jclepro.2017.01.081>
- Barreto, I. (2010). Dynamic Capabilities: A Review of Past Research and an Agenda for the

- Future. *Journal of Management*, 36(1), 256–280.
<https://doi.org/10.1177/0149206309350776>
- Bellman, R., Clark, C. E., Malcolm, D. G., Craft, C. J., & Ricciardi, F. M. (1957). On the Construction of a Multi-Stage, Multi-Person Business Game. *Operations Research*, 5(4), 469.
- Ben-Menahem, S. M., Kwee, Z., Volberda, H. W., & Van Den Bosch, F. A. J. (2013). Strategic renewal over time: The enabling role of potential absorptive capacity in aligning internal and external rates of change. *Long Range Planning*, 46(3), 216–235.
<https://doi.org/10.1016/j.lrp.2012.09.012>
- Bendle, N. T., & Wang, X. S. (2016). Uncovering the message from the mess of big data. *Business Horizons*, 59(1), 115–124. <https://doi.org/10.1016/j.bushor.2015.10.001>
- Bereznoi, A. (2014). Business Model Innovation in Corporate Competitive Strategy. *Problems of Economic Transition*, 57(8), 14–33. <https://doi.org/10.1080/10611991.2014.1042313>
- Bhatt, G. D., & Grover, V. (2005). Types of Information Technology Capabilities and Their Role in Competitive Advantage: An Empirical Study. *Journal of Management Information Systems*, 22(2), 253–277. <https://doi.org/10.1080/07421222.2005.11045844>
- Bhattacharya, M., Islam, R., & Abawajy, J. (2016). Evolutionary optimization: A big data perspective. *Journal of Network and Computer Applications*, 59, 416–426.
<https://doi.org/10.1016/j.jnca.2014.07.032>
- Bland, J. M., & Altman, D. G. (1997). Statistics Notes: Cronbach's Alpha. *British Medical Journal*, 314(7080), 572.
- Bommel, E. Van, Edelman, D., & Ungerman, K. (2014). Digitizing the consumer decision journey. *McKinsey&Company*, (June), 1–8.
- Boncea, R., Petre, I., Smada, D.-M., & Zamfiroiu Anagrama, A. (2017). A Maturity Analysis of Big Data Technologies. *Informatica Economica*, 21(1), 60–72.
<https://doi.org/10.12948/issn14531305/21.1.2017.05>
- Bourgeois, L. J., & Eisenhardt, K. M. (1988). Strategic decision processes in high velocity environments: Four cases in the microcomputer industry. *Management Science*, 34(7),

- 816–835. <https://doi.org/10.1287/mnsc.34.7.816>
- Boyd, D., & Crawford, K. (2012). Critical Questions for Big Data. *Information, Communication & Society*, 15(5), 37–41. <https://doi.org/10.1080/1369118X.2012.678878>
- Bradley, J., Barbier, J., & Handler, D. (2013). Embracing the Internet of Everything To Capture Your Share of \$ 14.4 Trillion. *Cisco Ibsg Group*, 2013. Retrieved from http://www.cisco.com/web/about/ac79/docs/innov/loE_Economy.pdf
- Braganza, A., Brooks, L., Nepelski, D., Ali, M., & Moro, R. (2017). Resource management in big data initiatives: Processes and dynamic capabilities. *Journal of Business Research*, 70, 328–337. <https://doi.org/10.1016/j.jbusres.2016.08.006>
- Brown, B., Chui, M., & Manyika, J. (2011). Are you ready for the era of Big Data? *Mckinsey Quarterly*, 1–12. <https://doi.org/00475394>
- Brown, & Moore, M. (2012). *Confirmatory Factor Analysis*. Boston, MA.
- Bryman, A., & Bell, E. (2007). *Business Research Methods*. New York: Oxford University Press. <https://doi.org/10.1016/B978-0-12-387000-1.01001-9>
- Brynjolfsson, E., Hitt, L. M., & Kim, H. H. (2011). Strength in Numbers: How does data-driven decision-making affect firm performance? *ICIS 2011 Proceedings*, 18. <https://doi.org/10.2139/ssrn.1819486>
- Buhl, H. U. (2013). Interview with Martin Petry on “Big Data.” *Business and Information Systems Engineering*, 5(2), 101–102. <https://doi.org/10.1007/s12599-013-0253-9>
- CEBMA. (2013). *Evidence-Based Management Assessment for Organizations*. Center for Evidence-Based Management.
- Chan, J. O. (2014). Big Data Customer Knowledge Management. *Communications of the IIMA*, 14(3), 45–56.
- Chen, C. L. P., & Zhang, C.-Y. (2014). Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. *Information Sciences*, 275, 314–347. <https://doi.org/10.1016/j.ins.2014.01.015>
- Chen, Chiang, R. H. L., & Storey, V. C. (2012). Business Intelligence and Analytics: From Big Data

- To Big Impact. *Mis Quarterly*, 36(4), 1165–1188.
<https://doi.org/10.1145/2463676.2463712>
- Chen, Mao, S., & Liu, Y. (2014). Big data: A survey. *Mobile Networks and Applications*, 19(2), 171–209. <https://doi.org/10.1007/s11036-013-0489-0>
- Cohen, W. M., & Levinthal, D. a. (1990). Absorptive capacity: A new perspective on learning and innovation. *Administrative Science Quarterly*, 35(1), 128–152.
<https://doi.org/10.2307/2393553>
- Cohen, W. M., & Levinthal, D. A. (1994). Fortune favors the prepared firm. *Management Science*, 40(2), 227–251. <https://doi.org/10.1287/mnsc.43.10.1455>
- Colas, M., Finck, I., Buvat, J., Nambiar, R., & Raj Singh, R. (2014). Cracking the Data Conundrum: How Successful Companies Make Big Data Operational. *Capgemini*, 1–17. Retrieved from https://www.capgemini-consulting.com/resource-file-access/resource/pdf/cracking_the_data_conundrum-big_data_pov_13-1-15_v2.pdf
- Comuzzi, M., & Patel, A. (2016). How Organisations Leverage Big Data: A Maturity Model. *Industrial Management & Data Systems*, 116(8), 31. <https://doi.org/10.1108/IMDS-12-2015-0495>
- Côrte-Real, N., Oliveira, T., & Ruivo, P. (2016). Assessing business value of Big Data Analytics in European firms. *Journal of Business Research*, 70, 379–390.
<https://doi.org/10.1016/j.jbusres.2016.08.011>
- 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(7), 1–9. <https://doi.org/10.1.1.110.9154>
- Coughlin, T. (2016). The cost of storage. Retrieved from <https://www.forbes.com/sites/tomcoughlin/2016/07/24/the-costs-of-storage/#2758f6b93239>
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297–334. <https://doi.org/10.1007/BF02310555>
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological*

- Bulletin*, 52(4), 281–302. <https://doi.org/10.1037/h0040957>
- 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. <https://doi.org/10.1111/j.1467-6486.2009.00880.x>
- Crowston, K., & Qin, J. (2011). A capability maturity model for scientific data management: Evidence from the literature. *Proceedings of the American Society for Information Science and Technology*, 48(1), 1–9. <https://doi.org/10.1002/meet.2011.14504801036>
- D'Aven, R. A. (1994). Hypercompetition - Managing the Dynamics of Strategic Maneuvering. In *Free Press* (p. 250).
- Damanpour, F. (1991). Organizational Innovation: a Meta-Analysis of Effects of Determinants and Moderators. *Academy of Management Journal*, 34(3), 555–590. <https://doi.org/10.2307/256406>
- Davenport, T. H., & Patil, D. J. (2011). Data Scientist: The Sexiest Job of the 21st Century. *Harvard Business Review*, (October 2012), 70–77. Retrieved from <https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century>
- Davison, M. L., Kwak, N., Seo, Y. S., & Choi, J. (2002). Using hierarchical linear models to examine moderator effects: Person-by-organization interactions. *Organizational Research Methods*, 5(3), 231–254. <https://doi.org/10.1177/1094428102005003003>
- 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. <https://doi.org/10.1007/s12599-014-0344-2>
- Domino Data Lab. (n.d.). Data Science Maturity Model. Retrieved from <https://www.dominodatalab.com/wp-content/uploads/2017/03/dsmm-final-v3.pdf>
- Donate, M. J., & Guadamillas, F. (2010). The effect of organizational culture on knowledge management practices and innovation. *Knowledge and Process Management*, 17(2), 82–94. <https://doi.org/10.1002/kpm.344>
- Dong, S., Johar, M. S., & Kumar, R. L. (2011). A Benchmarking Model for Management of Knowledge-Intensive Service Delivery Networks. *Journal of Management Information*

- Systems*, 28(3), 127–160. <https://doi.org/10.2753/MIS0742-1222280304>
- Drnevich, P. L., & Kriaucunas, A. P. (2011). clarifying the conditions and limits of the contributions of ordinary and dynamic capabilities to relative firm performance. *Strategic Management Journal*, 32, 254–279. <https://doi.org/10.1002/smj>
- Du, R. Y., Hu, Y., & Damangir, S. (2015). Leveraging Trends in Online Searches for Product Features in Market Response Modeling. *Journal of Marketing*, 79(1), 29–43. <https://doi.org/10.1509/jm.12.0459>
- Dutta, D., & Bose, I. (2015). Managing a Big Data project: The case of Ramco Cements Limited. *International Journal of Production Economics*, 165, 293–306. <https://doi.org/10.1016/j.ijpe.2014.12.032>
- Eisenhardt, K. M., & Martin, J. A. (2000). Dynamic Capabilities: What are they? *Strategic Management Journal*, 21, 1105–1121.
- El-Darwiche, B., Koch, V., Meer, D., & Tohme, D. W. (2014). Big data maturity: An action plan for policymakers and executives. *Strategy&*.
- Elbashir, M. Z., Collier, P. A., & Sutton, S. G. (2011). The Role of Organizational Absorptive Capacity in Strategic Use of Business Management Control Systems. *The Accounting Review*, 86(1), 155–184. <https://doi.org/10.2308/accr.00000010>
- Elgendy, N., & Elragal, A. (2016). Big Data Analytics in Support of the Decision Making Process. *Procedia Computer Science*, 100, 1071–1084. <https://doi.org/10.1016/j.procs.2016.09.251>
- Emani, C. K., Cullot, N., & Nicolle, C. (2015). ScienceDirect Understandable Big Data: A survey. *Computer Science Review*, 17, 70–81. <https://doi.org/10.1016/j.cosrev.2015.05.002>
- Erevelles, S., Fukawa, N., & Swayne, L. (2016). Big Data consumer analytics and the transformation of marketing. *Journal of Business Research*, 69(2), 897–904. <https://doi.org/10.1016/j.jbusres.2015.07.001>
- Espinosa, J. A., & Armour, F. (2016). The big data analytics gold rush: A research framework for coordination and governance. *Proceedings of the Annual Hawaii International Conference on System Sciences*, 2016–March, 1112–1121. <https://doi.org/10.1109/HICSS.2016.141>

- Etikan, I., Musa, S. A., & Alkassim, R. S. (2016). Comparison of Convenience Sampling and Purposive Sampling. *American Journal of Theoretical and Applied Statistics*, 5(1), 1. <https://doi.org/10.11648/j.ajtas.20160501.11>
- Eurich, M., Weiblen, T., & Breitenmoser, P. (2014). A six-step approach to business model innovation. *International Journal of Entrepreneurship and Innovation Management*, 18(4), 330–348. <https://doi.org/10.1504/IJEIM.2014.064213>
- Evans, J. R., & Mathur, A. (2005). The value of online surveys. *Internet Research*, 15(2), 195–219. <https://doi.org/10.1108/10662240510590360>
- Fan, S., Lau, R. Y. K., & Zhao, J. L. (2015). Demystifying Big Data Analytics for Business Intelligence Through the Lens of Marketing Mix. *Big Data Research*, 2(1), 28–32. <https://doi.org/10.1016/j.bdr.2015.02.006>
- Fang, K., Jiang, Y., & Song, M. (2016). Customer profitability forecasting using Big Data analytics: A case study of the insurance industry. *Computers & Industrial Engineering*, 101, 554–564. <https://doi.org/10.1016/j.cie.2016.09.011>
- Fichman, R. G. (2001). The Diffusion and Assimilation of Information Technology Innovations. *Framing the Domains of IT Management*, 105–127. <https://doi.org/10.1.1.24.4539>
- Fichman, R. G., & Kemerer, C. F. (1997). the Assimilation of Software Process Innovations : an Organizational Learning Perspective. *Management Science*, 43(10), 1345–1363. <https://doi.org/10.1287/mnsc.43.10.1345>
- Fong, N. M., Fang, Z., & Luo, X. (2015). Geo-Conquesting: Competitive Locational Targeting of Mobile Promotions. *Journal of Marketing Research*, 52(5), 726–735. <https://doi.org/10.1509/jmr.14.0229>
- Frambach, R. T., & Schillewaert, N. (2002). Organizational innovation adoption A multi-level framework of determinants and opportunities for future research. *Journal of Business Research*, 55, 163–176.
- Frankenberger, K., Weiblen, T., & Gassmann, O. (2014). The antecedents of open business models: an exploratory study of incumbent firms. *R&D Management Journal*, 44(2), 173–188. Retrieved from <https://www.alexandria.unisg.ch/Publikationen/nach-Institut/ITEM/230319>

- Fricker, R. D. (2008). Sampling Methods for Web and E-mail Surveys. In N. G. Fielding, R. M. Lee, & G. Blank (Eds.), *The SAGE Handbook of online research methods* (1st ed., pp. 195–216). London: SAGE Publications Ltd.
- 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. <https://doi.org/10.1016/j.ijinfomgt.2016.01.006>
- Gandomi, A., & Haider, M. (2015). the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, 35(2), 137–144. <https://doi.org/10.1016/j.ijinfomgt.2014.10.007>
- Gao, S., Yeoh, W., Wong, S. F., & Scheepers, R. (2017). A literature analysis of the use of Absorptive Capacity construct in IS research. *International Journal of Information Management*, 37(2), 36–42. <https://doi.org/10.1016/j.ijinfomgt.2016.11.001>
- Gassmann, O., Frankenberger, K., & Csik, M. (2013). The St . Gallen Business Model Navigator. *International Journal of Product Development*, 18(3), 249–273. https://doi.org/10.1007/978-3-319-01056-4_7
- Ghosh, K. (2016). Moving toward a Better Business Model for the Millennial Generation. *The Global Studies Journal*, 9(3). Retrieved from www.onglobalization.com
- Gil, D., & Song, I. Y. (2015). Modeling and Management of Big Data: Challenges and Opportunities. *Future Generation Computer Systems*, 63, 96–99. <https://doi.org/10.1016/j.future.2015.07.019>
- Giniuniene, J., & Jurksiene, L. (2015). Dynamic Capabilities, Innovation and Organizational Learning: Interrelations and Impact on Firm Performance. *Procedia - Social and Behavioral Sciences*, 213(1997), 985–991. <https://doi.org/10.1016/j.sbspro.2015.11.515>
- Glass, A. J., & Saggi, K. (1998). International technology transfer and the technology gap. *The Journal of Development Economics* VO - 55, 55(2), 369. [https://doi.org/10.1016/S0304-3878\(98\)00041-8](https://doi.org/10.1016/S0304-3878(98)00041-8)
- Gobble, M. M. (2013). Big Data: The Next Big Thing in Innovation. *Research Technology Management*, 56(1), 64–66. <https://doi.org/10.5437/08956308X5601005>

- Graham, J. W. (2009). Missing Data Analysis: Making It Work in the Real World. *Annual Review of Psychology*, 60(1), 549–576. <https://doi.org/10.1146/annurev.psych.58.110405.085530>
- Greenhalgh, T., Robert, G., Macfarlane, F., Bate, P., & Kyriakidou, O. (2004). Diffusion of Innovations in Service Organizations: Systematic Review and Recommendations. *Milbank Quarterly*, 82(4), 581–629.
- Guerra, N. G., & Knox, L. (2008). How culture impacts the dissemination and implementation of innovation: A case study of the Families and Schools Together program (FAST) for preventing violence with immigrant Latino youth. *American Journal of Community Psychology*, 41(3–4), 304–313. <https://doi.org/10.1007/s10464-008-9161-4>
- Gupta, M., & George, J. F. (2016). Toward the development of a big data analytics capability. *Information & Management*, 53(8), 1049–1064. <https://doi.org/10.1016/j.im.2016.07.004>
- Habib, M., Chang, V., Batool, A., & Ying, T. (2016). Big data reduction framework for value creation in sustainable enterprises. *International Journal of Information Management*, 36(6), 917–928. <https://doi.org/10.1016/j.ijinfomgt.2016.05.013>
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate Data Analysis* (Seventh Ed). Pearson. <https://doi.org/10.1016/j.ijpharm.2011.02.019>
- Halper, F., & Krishnan, K. (2014). TDWI Big Data Maturity Model Guide. *RDWi Research*, 2013–2014, 1–20.
- Harwell, M. R. (2011). Research Design in Qualitative/Quantitative/ Mixed Methods. In C. F. Conrad & R. C. Serlin (Eds.), *The SAGE Handbook for Research in Education: Pursuing Ideas as the Keystone of Exemplary Inquiry* (second, pp. 147–182). California: SAGE Publications, Inc. <https://doi.org/10.4135/9781412961288.n380>
- Hashem, I. A. T., Yaqoob, I., Anuar, N. B., Mokhtar, S., Gani, A., & Ullah Khan, S. (2015). The rise of “big data” on cloud computing: Review and open research issues. *Information Systems*, 47, 98–115. <https://doi.org/10.1016/j.is.2014.07.006>
- Helfat, C. E., & Peteraf, M. A. (2009). Understanding dynamic capabilities: preogress along a developmental path. *Strategic Organization*, 7(1), 91–102.

<https://doi.org/10.1177/1476127008100133>

- Hill, M. E. (2003). *The development of an instrument to measure intrapreneurship: entrepreneurship within the corporate setting*. Rhodes University.
- Hirsch, V. (2017). *AI & The Future of Work*. United Kingdom: TEDxManchester. Retrieved from <https://www.youtube.com/watch?v=dRw4d2Si8LA>
- Hofstede, G. (n.d.). Cultural Dimensions - Geert Hofstede. Retrieved from <https://geert-hofstede.com/cultural-dimensions.html>
- Honaker, J., & King, G. (2010). What to do about missing values in time-series cross-section data. *American Journal of Political Science*, 54(2), 561–581. <https://doi.org/10.1111/j.1540-5907.2010.00447.x>
- Horita, F. E. A., de Albuquerque, J. P., Marchezini, V., & Mendiondo, E. M. (2017). Bridging the gap between decision-making and emerging big data sources: an application of a model-based framework to disaster management in Brazil. *Decision Support Systems*, 97, 12–22. <https://doi.org/10.1016/j.dss.2017.03.001>
- Hortonworks. (2016). Hortonworks big data maturity model. *Hortonworks*, (March).
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6(1), 1–55. <https://doi.org/10.1080/10705519909540118>
- Huang, G. Q., Zhong, R. Y., & Tsui, K. L. (2015). Special issue on “Big data for service and manufacturing supply chain management.” *International Journal of Production Economics*, 165, 172–173. <https://doi.org/10.1016/j.ijpe.2015.05.009>
- Jackson, D. L., Gillaspay, J. A., & Purc-Stephenson, R. (2009). Reporting practices in confirmatory factor analysis: An overview and some recommendations. *Psychological Methods*, 14(1), 6–23. <https://doi.org/10.1037/a0014694>
- Jagadish, H. V. (2015). Big Data and Science: Myths and Reality. *Big Data Research*, 2(2), 49–52. <https://doi.org/10.1016/j.bdr.2015.01.005>
- Janssen, M., Voort, H. Van Der, & Wahyudi, A. (2017). Factors influencing big data decision-making quality. *Journal of Business Research*, 70, 338–345.

<https://doi.org/10.1016/j.jbusres.2016.08.007>

- Jin, X., Wah, B. W., Cheng, X., & Wang, Y. (2015). Significance and Challenges of Big Data Research. *Big Data Research*, 2(2), 59–64. <https://doi.org/10.1016/j.bdr.2015.01.006>
- Johnson, R. B. (1997). Examining the validity structure of qualitative research. *Education*, 118(2), 282–292. <https://doi.org/Retrieved> from <http://www.nova.edu/ssss/QR/QR4-3/winter.html>
- Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika*, 39, 31–36.
- Kaisler, S., Armour, F., Espinosa, J. A., & Money, W. (2013). Big Data: Issues and Challenges Moving Forward. *2013 46th Hawaii International Conference on System Sciences*, 995–1004. <https://doi.org/10.1109/HICSS.2013.645>
- Kambatla, K., Kollias, G., Kumar, V., & Grama, A. (2014). Trends in big data analytics. *Journal of Parallel and Distributed Computing*, 74(7), 2561–2573. <https://doi.org/10.1016/j.jpdc.2014.01.003>
- Kane, G. C., Palmer, D., Phillips, A., & Kiron, D. (2015). Is Your Business Ready for a Digital Future? *MIT Sloan Management Review*, 56(4), 37–44.
- Kane, G. C., Palmer, D., Phillips, A. N., Kiron, D., & Buckley, N. (2016). *Aligning the Organization for Its Digital Future*. MITSloan Management Review and Deloitte University Press.
- Keller, W. (1996). Absorptive capacity: On the creation and acquisition of technology in development. *Journal of Development Economics*, 49(1), 199–227. [https://doi.org/10.1016/0304-3878\(95\)00060-7](https://doi.org/10.1016/0304-3878(95)00060-7)
- Khan, S., Liu, X., Shakil, K. A., & Alam, M. (2017). A survey on scholarly data: From big data perspective. *Information Processing & Management*, 53(4), 923–944. <https://doi.org/10.1016/j.ipm.2017.03.006>
- Kiesler, S., & Sproull, L. S. (1986). Response effects in the electronic survey. *Public Opinion Quarterly*. <https://doi.org/10.1086/268992>
- Kim, G., Shin, B., & Kwon, O. (2012). Investigating the Value of Sociomaterialism in Conceptualizing IT Capability of a Firm. *Journal of Management Information Systems*, 29(3), 327–362. <https://doi.org/10.2753/MIS0742-1222290310>

- King, G., Honaker, J., Joseph, A., & Scheve, K. (2001). Analyzing Incomplete Political Science Data: An Alternative Algorithm for Multiple Imputation. *American Political Science Review*, 85(1269), 49–69. <https://doi.org/10.2307/3117628>
- Kiron, D., Prentice, P. K., & Ferguson, R. B. (2014). The Analytics Mandate. *MIT Sloan Management Review*, 55(4), 1.
- Kiron, D., & Shockley, R. (2011). Creating Business Value with Analytics. *MIT Sloan Management Review*, 53(1), 57–63.
- Kline, R. B. (2016). *Principles and practice of structural equation modeling* (Fourth Edi). New York: Guilford Press.
- Knofczynski, G. T., & Mundfrom, D. (2008). Sample Sizes When Using Multiple Linear Regression for Prediction. *Educational and Psychological Measurement*, 68(3), 431–442. <https://doi.org/10.1177/0013164407310131>
- Komorowski, M. (2009). A history of storage cost. Retrieved from <http://www.mkomo.com/cost-per-gigabyte>
- Kosieradzka, A. (2017). Maturity Model for Production Management. *Procedia Engineering*, 182, 342–349. <https://doi.org/http://doi.org/10.1016/j.proeng.2017.03.109>
- Kshetri, N. (2014). Big datas impact on privacy, security and consumer welfare. *Telecommunications Policy*, 38(11), 1134–1145. <https://doi.org/10.1016/j.telpol.2014.10.002>
- Kubina, M., Varmus, M., & Kubinova, I. (2015). Use of big data for competitive advantage of company. *Procedia Economics and Finance*, 26(15), 561–565. [https://doi.org/10.1016/S2212-5671\(15\)00955-7](https://doi.org/10.1016/S2212-5671(15)00955-7)
- Kuratko, D. F., Montagno, R. V., & Hornsby, J. S. (1990). Developing an Intrapreneurial Assessment Instrument for an Effective Corporate Entrepreneurial. *Strategic Management Journal*, 11, 49–58.
- LaChapelle, C. (2016). The cost of data storage and management: Where is it headed in 2016? *The Data Center Journal*, March. Retrieved from <http://www.datacenterjournal.com/cost-data-storage-management-headed-2016/>

- Lai, P. C. (2017). The literature review of technology adoption models and theories for the novelty technology. *Journal of Information Systems and Technology Management*, 14(1), 21–38. <https://doi.org/10.4301/S1807-17752017000100002>
- Lane, P. J., & Lubatkin, M. (1998). Relative Absorptive Capacity and Interorganizational Learning. *Strategic Management Journal*, 19(5), 461–477.
- Laroche, M., Bergeron, J., & Barbaro-Forleo, G. (2001). Targeting consumers who are willing to pay more for environmentally friendly products. *Journal of Consumer Marketing*, 18(6), 503–520.
- Larson, D., & Chang, V. (2016). A review and future direction of agile, business intelligence, analytics and data science. *International Journal of Information Management*, 36(5), 700–710. <https://doi.org/10.1016/j.ijinfomgt.2016.04.013>
- Lau, L., Yang-turner, F., & Karacapilidis, N. (2014). Requirements for Big Data Analytics Supporting Decision Making: A Sensemaking Perspective. In *Mastering Data-Intensive Collaboration and Decision Making* (Vol. 5, pp. 49–70). <https://doi.org/10.1007/978-3-319-02612-1>
- 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.
- Lee, I. (2017). Big data: Dimensions, evolution, impacts, and challenges. *Business Horizons*. <https://doi.org/10.1016/j.bushor.2017.01.004>
- Leidner, D. E., & Kayworth, T. (2006). A review of culture in Information systems research: Toward a Theory of Information Technology Culture confluct. *Mis Quarterly*, 30(2), 357–399.
- Lismont, J., Vanthienen, J., Baesens, B., & Lemahieu, W. (2017). Defining analytics maturity indicators: A survey approach. *International Journal of Information Management*, 37(3), 114–124. <https://doi.org/10.1016/j.ijinfomgt.2016.12.003>
- Liu, J., Li, J., Li, W., & Wu, J. (2016). Rethinking big data: A review on the data quality and usage issues. *ISPRS Journal of Photogrammetry and Remote Sensing*, 115, 134–142. <https://doi.org/10.1016/j.isprsjprs.2015.11.006>

- Loebbecke, C., & Picot, A. (2015). Reflections on societal and business model transformation arising from digitization and big data analytics: A research agenda. *Journal of Strategic Information Systems*, 24(3), 149–157. <https://doi.org/10.1016/j.jsis.2015.08.002>
- MacCallum, R. C., & Austin, J. T. (2000). Applications of Structural Equation Modelling in Psychological research. *Annual Reviews of Psychology*, 51(August), 201–226.
- Malthouse, E. C., Haenlein, M., Skiera, B., Wege, E., & Zhang, M. (2013). Managing customer relationships in the social media era: Introducing the social CRM house. *Journal of Interactive Marketing*, 27(4), 270–280. <https://doi.org/10.1016/j.intmar.2013.09.008>
- Manfreda, A., Kovacic, A., Štemberger, M. I., & Trkman, P. (2014). Absorptive capacity as a precondition for business process improvement. *Journal of Computer Information Systems*, 54(2), 35–43. <https://doi.org/10.1080/08874417.2014.11645684>
- Marsh, H. W., Hau, K.-T., & Wen, Z. (2004). In Search of Golden Rules Comment on Hypothesis Testing Approaches to Setting Cutoff Values for Fit Indexes and Dangers in Overgeneralizing Hu and Bentler's (1999) Findings. *Structural Equation Modeling*, 11(3), 320–341. <https://doi.org/10.1207/s15328007sem1103>
- Martín-de Castro, G., Delgado-Verde, M., Navas-López, J. E., & Cruz-González, J. (2013). The moderating role of innovation culture in the relationship between knowledge assets and product innovation. *Technological Forecasting and Social Change*, 80(2), 351–363. <https://doi.org/10.1016/j.techfore.2012.08.012>
- Mason, J. (2002). *Qualitative Researching* (Second Edi). London: SAGE Publications Ltd.
- Maxwell, S. E., & Delaney, H. D. (2004). *Designing experiments and analyzing data: A model comparison perspective* (Vol. 1). Psychology Press.
- Mayer, K. J., & Teece, D. J. (2008). Unpacking strategic alliances: The structure and purpose of alliance versus supplier relationships. *Journal of Economic Behavior and Organization*, 66(1), 106–127. <https://doi.org/10.1016/j.jebo.2007.06.007>
- Mazzei, M. J., & Noble, D. (2017). Big data dreams: A framework for corporate strategy. *Business Horizons*. <https://doi.org/10.1016/j.bushor.2017.01.010>
- McAbee, S. T., Landis, R. S., & Burke, M. I. (2016). Inductive reasoning: The promise of big data.

- Human Resource Management Review*, 27(2), 277–290.
<https://doi.org/10.1016/j.hrmr.2016.08.005>
- McAfee, A., & Brynjolfsson, E. (2012). Big data: the management revolution. *Harvard Business Review*, 90(10). <https://doi.org/10.1007/s12599-013-0249-5>
- Mettler, T., Winter, R., Mettler, T., Rohner, P., & Winter, R. (2010). Management of the Interconnected World. In A. D'Atri, M. De Marco, F. Cabiddu, & M. Braccini (Eds.), *Management of the Interconnected World* (pp. 333–340). Berlin Heidelberg: Springer-Verlag. <https://doi.org/10.1007/978-3-7908-2404-9>
- Mikalef, P., & Pateli, A. (2017). Information technology-enabled dynamic capabilities and their indirect effect on competitive performance: Findings from PLS-SEM and fsQCA. *Journal of Business Research*, 70, 1–16. <https://doi.org/10.1016/j.jbusres.2016.09.004>
- Miloslavskaya, N., & Tolstoy, A. (2016). Big Data, Fast Data and Data Lake Concepts. *Procedia - Procedia Computer Science*, 88, 300–305. <https://doi.org/10.1016/j.procs.2016.07.439>
- Monino, J.-L. (2016). Data Value, Big Data Analytics, and Decision-Making. *Journal of the Knowledge Economy*, (July). <https://doi.org/10.1007/s13132-016-0396-2>
- Nair, H. S., Misra, S., Hornbuckle, W. J., Mishra, R., & Acharya, A. (2014). Big Data and Marketing Analytics in Gaming: Combining Empirical Models and Field Experimentation. *Working Papers (Faculty) -- Stanford Graduate School of Business*, 1–38. Retrieved from <http://ezproxy.unicartagena.edu.co:2048/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=98738743&lang=es&site=ehost-live>
- Nannetti, P. (2012). The Deciding Factor: Big Data & Decision Making. *Capgemini*.
- Oliveira, T., & Martins, M. F. (2011). Literature Review of Information Technology Adoption Models at Firm Level. *The Electronic Journal Information Systems Evaluation*, 14(1), 110–121.
- Oliveira, T., Thomas, M. A., & Espadanal, M. (2014). Assessing the Determinants of Cloud Computing Adoption: An Analysis of the Manufacturing and ... *Information & Management*, 51(July), 497–510. <https://doi.org/10.1016/j.im.2014.03.006>
- Orlikowski, W., & Baroudi, J. J. (1990). Studying Information Technology in organisations:

- Research approaches and assumptions. *Center for Digital Economy Research Stem School of Business*, 1–31. <https://doi.org/10.1287/isre.2.1.1>
- Ostrom, A. L., Parasuraman, A., Bowen, D., Patrício, L., & Voss, C. A. (2015). Service Research Priorities in a Rapidly Changing Context. *Journal of Service Research*, 18(2), 127–159.
- Özköse, H., Arı, E. S., & Gencer, C. (2015). Yesterday, Today and Tomorrow of Big Data. *Procedia - Social and Behavioral Sciences*, 195, 1042–1050. <https://doi.org/10.1016/j.sbspro.2015.06.147>
- Paulk, M. C., Curtis, B., Chrissis, M. B., & Weber, C. V. (1993). Capability Maturity Model for Softwar, Version 1.1. *Carnegie Melon University*, (February), 82.
- 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. <https://doi.org/10.1287/isre.1060.0094>
- Perez-Soltero, A., & Soto, V. L. (2017). A Model Based on Core Processes and Knowledge Management to Promote Innovation: A Case of a Mexican Trading Company. *IUP Journal of Knowledge Management* 15, 7–29.
- Perrons, R. K., & Jensen, J. W. (2015). Data as an asset: What the oil and gas sector can learn from other industries about “Big Data.” *Energy Policy*, 81, 117–121. <https://doi.org/10.1016/j.enpol.2015.02.020>
- Pisano, G. P., & Teece, D. J. (2007). How to Capture Value from Innovation: Shaping Intellectual Property and Industry Architecture. *California Management Review*, 50(1), 278–296. <https://doi.org/Article>
- 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. <https://doi.org/10.1016/j.dss.2012.08.017>
- Pospiech, M., & Felden, C. (2012). Big data—a state-of-the-art. *AMCIS 2012 Proceedings*, 1–11. Retrieved from <http://aisel.aisnet.org/amcis2012/proceedings/DecisionSupport/22/>
- Power, D. J. (2015). “Big Data” Decision Making Use Cases. *Springer International Publishing*, 1–9. <https://doi.org/10.1007/978-3-319-18533-0>

- Provost, F., & Fawcett, T. (2013). Data Science and its Relationship to Big Data and Data-Driven Decision Making. *Big Data*, 1(1), 51–59. <https://doi.org/10.1089/big.2013.1508>
- Rad, M. S., Nilashi, M., & Dahlan, H. M. (2017). Information technology adoption: A review of the literature and classification. *Universal Access in the Information Society*, 1(30). <https://doi.org/10.1007/s10209-017-0534-z>
- Rajterič, I. H. (2010). Overview of business intelligence maturity models. *Management: Journal of Contemporary Management ...*, 15, 47–67. <https://doi.org/10.1007/BF01530759>
- Ram, J., Zhang, C., & Koronios, A. (2016). The implications of Big Data analytics on Business Intelligence: A qualitative study in China. *Procedia - Procedia Computer Science*, 87, 221–226. <https://doi.org/10.1016/j.procs.2016.05.152>
- Ravichandran, T., & Lertwongsatien, C. (2005). Effect of Information Systems Resources and Capabilities on Firm Performance: A Resource-Based Perspective. *Journal of Management Information Systems*, 21(4), 237–276.
- Razali, N. M., & Wah, Y. B. (2011). Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. *Journal of Statistical Modeling and Analytics*, 2(1), 21–33. <https://doi.org/doi:10.1515/bile-2015-0008>
- 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.
- Roelens, B., Steenacker, W., & Poels, G. (2017). Realizing strategic fit within the business architecture: the design of a Process-Goal Alignment modeling and analysis technique. *Software & Systems Modeling*. <https://doi.org/10.1007/s10270-016-0574-5>
- Rogers, P., & Meehan, P. (2007). Building a winning culture. *Bain & Company, Inc*, 8(4), 254–261. <https://doi.org/10.1108/17515630710684420>
- Rubin, D. B. (1976). Inference and Missing Data. *Biometrika*. <https://doi.org/10.2307/2335739>
- Saunders, M., & Lewis, P. (2012). *Doing Research in Business & Management. An Essential Guide to Planning Your Project*. Harlow: Pearson Education Limited.
- Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art.

- Psychological Methods*, 7(2), 147–177. <https://doi.org/10.1037/1082-989X.7.2.147>
- Schaller, R. R. (1997). Moore's law: past, present and future. *Spectrum, IEEE*, 34(6), 52–59. <https://doi.org/10.1109/6.591665>
- Scheffer, J. (2002). Dealing With Missing Data. *Research Letters in the Information and Mathematical Sciences*, 3, 153–160. <https://doi.org/10.1016/j.pmrj.2015.07.011>
- Schilke, O. (2014). On the contingent value of dynamic capabilities for competitive advantage: The nonlinear moderating effect of environmental dynamism. *Strategic Management Journal*, 35, 197–203. <https://doi.org/10.1002/smj>
- Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., King, J., Nora, A., & Barlow, E. A. (2006). Reportig Structural Equation Modeling and Confirmatory Factor Analysis Results: A Review. *The Journal of Educational Research*, 99(6), 232–338. <https://doi.org/10.3200/JOER.99.6.323-338>
- Shah, N., Irani, Z., Sharif, A. M., & Job, A. (2017). Big data in an HR context: Exploring organizational change readiness, employee attitudes and behaviors Readiness. *Journal of Business Research*, 70, 366–378. <https://doi.org/10.1016/j.jbusres.2016.08.010>
- Shahzad, F., Xiu, G., & Shahbaz, M. (2017). Organizational culture and innovation performance in Pakistan's software industry. *Technology in Society*, 51, 66–73. <https://doi.org/10.1016/j.techsoc.2017.08.002>
- Shapiro, A. S. S., & Wilk, M. B. (1965). An Analysis of Variance Test for Normality (Complete Samples). *Biometrika*, 52(3), 591–611.
- Sher, P. J., & Lee, V. C. (2004). Information Technology as a Facilitator of Enhancing Dynamic Capability through Knowledge Management. *Infofrmation & Management*, 41(8), 933–945.
- Shin, D. (2016). Demystifying big data: Anatomy of big data developmental process. *Telecommunications Policy*, 40(9), 837–854. <https://doi.org/10.1016/j.telpol.2015.03.007>
- Siddiq, A., Hashem, I. A. T., Yaqoob, I., Marjani, M., Shamshirband, S., Gani, A., & Nasaruddin, F. (2016). A survey of big data management: Taxonomy and state-of-the-art. *Journal of Network and Computer Applications*, 71, 151–166.

<https://doi.org/10.1016/j.jnca.2016.04.008>

- Sil, J., & Das, A. K. (2016). Feature Selection for Adaptive Decision Making in Big Data Analytics. In *Data Science and Big Data Computing* (Vol. 58, pp. 7250–7). <https://doi.org/10.1007/978-3-319-31861-5>
- Simonite, T. (2016). Moore's Law is Dead. Now What? *MIT Technology Review*, May 13, 40–41. Retrieved from <https://www.technologyreview.com/s/601441/moores-law-is-dead-now-what/>
- Singh, D., & Reddy, C. K. (2014). A survey on platforms for big data analytics. *Journal of Big Data*, 2(1), 8. <https://doi.org/10.1186/s40537-014-0008-6>
- Sivarajah, U., Kamal, M. M., Irani, Z., & Weerakkody, V. (2017). Critical analysis of Big Data challenges and analytical methods. *Journal of Business Research*, 70, 263–286. <https://doi.org/10.1016/j.jbusres.2016.08.001>
- Štefan, S., & Branislav, Z. (2016). Relationship between Business Strategy and Business Model Studied in a Sample of Service Companies, 8(4), 72–84. <https://doi.org/10.7441/joc.2016.04.05>
- Straub, D., Boudreau, M., & Gefen, D. (2004). Validation guidelines for IS positivist research Validation Guidelines for IS Positivist Research. *Communications of the Association for Information Systems*, 13(January), 380–427. <https://doi.org/Article>
- Summers, C. A., Smith, R. W., & Reczek, R. W. (2016). An audience of one: Behaviorally targeted ads as implied social labels. *Journal of Consumer Research*, 43(1), 156–178. <https://doi.org/10.1093/jcr/ucw012>
- Tavallaei, R., Shokohyar, S., Moosavi, S. M., & Sarfi, Z. (2015). Assessing the Evaluation Models of Business Intelligence Maturity and Presenting an Optimized Model. *International Journal of Management, Accounting and Economics*, 2(9), 1005–1019. Retrieved from www.ijmae.com
- Teece, D. J. (2007). Explicating Dynamic Capabilities: The Nature and Microfoundations of (Long-Run) Enterprise Performance. *Working Paper*, 28(13), 23.
- Teece, D. J. (2010). Business models, business strategy and innovation. *Long Range Planning*.

- Teece, D. J., & Leih, S. (2016). Uncertainty, Innovation, and Dynamic Capabilities: an Introduction. *California Management Review*, 58(4), 5–13.
- Teece, D. J., Peteraf, M., & Leih, S. (2016). Dynamic Capabilities and Organizational Agility: Risk, Uncertainty, and Strategy in the Innovation Economy. *California Management Review*, 58(4), 13–36. <https://doi.org/10.5277/ord130406>
- Teece, D. J., Pisano, G. P., & Shuen, A. (1997). Dynamic capabilities and strategic management. *Strategic Management Journal*, 18(7), 509–533. [https://doi.org/10.1002/\(SICI\)1097-0266\(199708\)18:7<509::AID-SMJ882>3.0.CO;2-Z](https://doi.org/10.1002/(SICI)1097-0266(199708)18:7<509::AID-SMJ882>3.0.CO;2-Z)
- Teece, D. J., Rumelt, R., Dosi, G., & Winter, S. (1994). Theory corporate coherence and evidence. *Journal of Economic Behavior and Organisation*, 23, 1–30.
- Van den Bosch, F. A. J., Volberda, H. W., & de Boer, M. (1999). Coevolution of Firm Absorptive Capacity and Knowledge Environment: Organizational Forms and Combinative Capabilities. *Organization Science*, 10(5), 551–568. <https://doi.org/10.1287/orsc.10.5.551>
- Vargo, S. L., & Lusch, R. F. (2004). Evolving to a New Dominant Logic for Marketing. *Journal of Marketing*, 68, 1–17. <https://doi.org/10.1509/jmkg.68.1.1.24036>
- Von Den Driesch, T., Eva, M., Flatten, T. C., & Brettel, M. (2015). How CEO experience, personality, and network affect firms' dynamic capabilities. *European Management Journal*, 33(4), 245–256. <https://doi.org/10.1016/j.emj.2015.01.003>
- Waller, M. A., & Fawcett, S. E. (2013). 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. <https://doi.org/10.1111/jbl.12010>
- 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. <https://doi.org/10.1016/j.ijpe.2014.12.031>
- 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*, 70, 356–365. <https://doi.org/10.1016/j.jbusres.2016.08.009>

- Wang, C. L., & Ahmed, P. K. (2007). Dynamic capabilities: A review and research agenda. *International Journal of Management Reviews*, 9(1), 31–51. <https://doi.org/10.1111/j.1468-2370.2007.00201.x>
- Wang, N., Liang, H., & Zhong, W. (2012). Resource Structuring or Capability Building? An Empirical Study of the Business Value of Information Technology. *Journal of Management Information Systems*, 29(2), 325–367. <https://doi.org/10.2753/MIS0742-1222290211>
- Wang, Y., & Hajli, N. (2017). Exploring the path to big data analytics success in healthcare. *Journal of Business Research*, 70, 287–299. <https://doi.org/10.1016/j.jbusres.2016.08.002>
- Wang, Y., Kung, L. A., & Byrd, T. A. (2015). Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations. *Technological Forecasting and Social Change*. <https://doi.org/10.1016/j.techfore.2015.12.019>
- Wang, Xu, Z., Fujita, H., & Liu, S. (2016). Towards felicitous decision making: An overview on challenges and trends of Big Data. *Information Sciences*, 367–368, 747–765. <https://doi.org/10.1016/j.ins.2016.07.007>
- Weiner, B. J. (2009). A theory of organizational readiness for change. *Implementation Science*, 4(1), 67–76. <https://doi.org/10.1186/1748-5908-4-67>
- Williams, B., Onsman, A., & Brown, T. (1996). Exploratory factor analysis: A five-step guide for novices. *Journal of Emergency Primary Health Care*, 19(May), 42–50. <https://doi.org/10.1080/09585190701763982>
- Williams, S. (2003). The business value of business intelligence. *Business Intelligence Journal*, 8(301), 30–39. Retrieved from http://www.decisionpath.com:8180/DPC_revised/html/BIJarticle.pdf%5Cnhttp://kingdomcreationscenter.org/wp-content/uploads/2010/08/the-business-value-of-business-intelligence1.pdf
- Winter, G. (2000). The Qualitative Report A Comparative Discussion of the Notion of “Validity” in Qualitative and Quantitative Research. *The Qualitative Report*, 4(3), 1–14. [https://doi.org/Retrieved from http://www.nova.edu/ssss/QR/QR4-3/winter.html](https://doi.org/Retrieved%20from%20http://www.nova.edu/ssss/QR/QR4-3/winter.html)
- Wisdom, J. P., Chor, K. H. B., Hoagwood, K. E., & Horwitz, S. M. (2014). Innovation Adoption: A Review of Theories and Constructs. *Administration and Policy in Mental Health Services*

- Research*, 41, 480–502. <https://doi.org/10.1007/s10488-013-0486-4>
- Wrigley, C., Bucolo, S., & Straker, K. (2016). Designing new business models: blue sky thinking and testing. *Journal of Business Strategy*, 37(5), 22–31. <https://doi.org/10.1108/JBS-04-2015-0041>
- Xu, Z., Frankwick, G. L., & Ramirez, E. (2016). Effects of big data analytics and traditional marketing analytics on new product success: A knowledge fusion perspective. *Journal of Business Research*, 69(5), 1562–1566. <https://doi.org/10.1016/j.jbusres.2015.10.017>
- Yaqoob, I., Hashem, I. A. T., Gani, A., Mokhtar, S., Ahmed, E., Anuar, N. B., & Vasilakos, A. V. (2016). Big data: From beginning to future. *International Journal of Information Management*, 36(6), 1231–1247. <https://doi.org/10.1016/j.ijinfomgt.2016.07.009>
- Zahra, S. A., & George, G. (2002). Absorptive Capacity: a Review, Reconceptualisation, and Extension. *Academy of Management Review*, 27(2), 185–203. <https://doi.org/10.5465/APBPP.2000.5438568>
- Zaki, M. J., & Meira, M. J. (2013). *Data Mining and Analysis: Fundamental Concepts and Algorithms*. Cambridge University Press.
- 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*, 101, 572–591. <https://doi.org/10.1016/j.cie.2016.07.013>
- Zhou, L., Pan, S., Wang, J., & Vasilakos, A. V. (2017). Machine learning on big data: Opportunities and challenges. *Neurocomputing*, 237(September 2016), 350–361. <https://doi.org/10.1016/j.neucom.2017.01.026>
- Zikmund, W. G., Babin, B. J., Carr, J. C., & Griffin, M. (2012). *Business Research Methods* (Eighth Edi). Mason: South-Western, Cengage Learning.
- Zuiderveen Borgesius, F. J. (2016). Singling out people without knowing their names - Behavioural targeting, pseudonymous data, and the new Data Protection Regulation. *Computer Law and Security Review*, 32(2), 256–271. <https://doi.org/10.1016/j.clsr.2015.12.013>

Appendices

9.1. Appendix 1: Research Questionnaire

9.1.1. Introduction

We are conducting research on how firms should implement a big data analytics capability within a business infrastructure and what business capabilities are necessary in order to extract sustained value. We contrast big data analytics, as a capability that must be developed in a business, with organisational characteristics including organisational culture to big data analytics capability, data driven decision making maturity and the dynamic capability of the firm (ability to adapt to changes in the business environment). To this end, you are asked to complete a short survey should take no more than 15 minutes.

In addition to all data being collected anonymously and confidentially, your participation is voluntary and you can withdraw at any time. By completing the survey, you indicate that you voluntarily participate in this research.

If you have any concerns, please contact myself or my research supervisor with the following details;

Researcher name: Dr Michael Niland

Email: michaeln@blts.co.za

Phone: 082 891 4142

Research Supervisor: Manoj Chiba

Email: ChibaM@gibs.co.za

9.1.2. Questionnaire

Instructions: Please select the most appropriate answer according to the scale provided in each question below;

Table 21: Survey Questionnaire and Answer Options

Section 1: Context of Organisation and Respondent								
1	Are you aware of or associated with a Big Data Analytics Capability within the organisation being described in this questionnaire?	Yes	No					
2	Which of the following best describes the principal industry of your organisation?	Standard SurveyMonkey™ drop down menu of industries						
3	How long have you been associated with the field of big data or data analytics?	0-2	2-5	5-10	10-15	15+		
4	How long has your organisation actively pursued or applied big data analytics to its business?	0-2	2-4	4-6	6-8	8+		
5	What is your main association with the data analytics capability?	User of analytics within business	Data analyst (Direct processor of data)	IT Systems or Infrastructure (Data technology environment)	Big Data Management (Driving application of resources)	Other (please Specify)		
6	What is the approximate total number of employees within your organisation?	1-99	100-499	500-999	1000 or more	Don't know		
7	Which of the following best describes your current job level?	Owner/ Executive/ C-Level	Senior Management	Middle Management	Intermediate	Entry Level	Other (please specify)	
8	In what country do you work?	Standard SurveyMonkey™ drop down menu of countries						
Section 2: Organisational Culture to Big Data Analytics Capability								
9	Our organisation has a widely held belief that innovation is an absolute necessity for the organisation's future	Strongly Disagree	Disagree	Moderately Disagree	Indifferent	Moderately Agree	Agree	Strongly Agree
10	Our organisation enables learning, accumulation and application of new knowledge better than our competitors	Strongly Disagree	Disagree	Moderately Disagree	Indifferent	Moderately Agree	Agree	Strongly Agree
11	We believe it is important to adopt new and cutting-edge practices to continuously improve product or service delivery	Strongly Disagree	Disagree	Moderately Disagree	Indifferent	Moderately Agree	Agree	Strongly Agree
12	People in our organisation are continuously encouraged to expand their capacities to achieve more and apply new capabilities	Never	Very Rarely	Rarely	Occasionally	Frequently	Very Frequently	Always
13	Our organisation can be described as visionary and flexible	Strongly Disagree	Disagree	Moderately Disagree	Indifferent	Moderately Agree	Agree	Strongly Agree

14	There is an extensive employee orientation program for new employees to ensure employees share the corporate vision and purpose	Strongly Disagree	Disagree	Moderately Disagree	Indifferent	Moderately Agree	Agree	Strongly Agree
15	We invest in targeted training and support at all levels of our organisation to assist our organisation to understand or know how to use data that is available	Strongly Disagree	Disagree	Moderately Disagree	Indifferent	Moderately Agree	Agree	Strongly Agree
16	Our executive level actively and visibly support our big data analytics capability	Strongly Disagree	Disagree	Moderately Disagree	Indifferent	Moderately Agree	Agree	Strongly Agree
Section 3: Data Driven Decision Making in the Organisation								
17	Before any decision is taken we systematically evaluate internal data to better understand the nature of the problem and what to do about it	Never	Very Rarely	Rarely	Occasionally	Frequently	Very Frequently	Always
18	Employees receive recognition from the organisation for applying evidence based decision making in our typical business processes	Never	Very Rarely	Rarely	Occasionally	Frequently	Very Frequently	Always
19	Managers in our organisation tend to believe that experience and knowledge gained on the job is the only important source of information when considering how to tackle a problem	Strongly Disagree	Disagree	Moderately Disagree	Indifferent	Moderately Agree	Agree	Strongly Agree
20	We make decisions by looking at what other organisations are doing, and how it's working for them	Never	Very Rarely	Rarely	Occasionally	Frequently	Very Frequently	Always
21	Internal politics and power struggles influence the way we make decisions about policies and practices	Never	Very Rarely	Rarely	Occasionally	Frequently	Very Frequently	Always
22	We believe that our competitiveness depends on our analytics capability	Strongly Disagree	Disagree	Moderately Disagree	Indifferent	Moderately Agree	Agree	Strongly Agree
23	Our decision makers all have access to a management information system	Never	Very Rarely	Rarely	Occasionally	Frequently	Very Frequently	Always
24	Our managers know how to critically appraise both internal data and evidence from scientific research	None of them			Some of them			All of them
Section 4: Dynamic Capability / Adsorptive Capacity of the Organisation								
25	Our organisation can be described as a bureaucratic organisation	Strongly Disagree	Disagree	Moderately Disagree	Indifferent	Moderately Agree	Agree	Strongly Agree
26	Our organisation's structure allows for resource sharing and encourages flexibility	Strongly Disagree	Disagree	Moderately Disagree	Indifferent	Moderately Agree	Agree	Strongly Agree
27	Our organisation has introduced more products, services and/or business capabilities than our competitors in the past two years	Strongly Disagree	Disagree	Moderately Disagree	Indifferent	Moderately Agree	Agree	Strongly Agree

28	Our business processes are flexible to allow for quick infusion of new methodologies, tools, techniques, insight or capabilities	Strongly Disagree	Disagree	Moderately Disagree	Indifferent	Moderately Agree	Agree	Strongly Agree
29	Our organisation has better communication and coordination than our competitors	Strongly Disagree	Disagree	Moderately Disagree	Indifferent	Moderately Agree	Agree	Strongly Agree
30	The product/services technologies change very quickly in our industry	Strongly Disagree	Disagree	Moderately Disagree	Indifferent	Moderately Agree	Agree	Strongly Agree
31	Our organisation has better capability of resource deployment than our competitors	Strongly Disagree	Disagree	Moderately Disagree	Indifferent	Moderately Agree	Agree	Strongly Agree
Section 5: Big Data Analytics Management Capabilities								
32	We continuously examine the innovative opportunities for the strategic use of big data analytics	Strongly Disagree	Disagree	Moderately Disagree	Indifferent	Moderately Agree	Agree	Strongly Agree
33	When we make business analytics investment decisions, we estimate the time managers will need to spend overseeing the change	Never	Very Rarely	Rarely	Occasionally	Frequently	Very Frequently	Always
34	When we make big data analytics investment decisions, we project about how much these options will help end-users make quicker decisions	Never	Very Rarely	Rarely	Occasionally	Frequently	Very Frequently	Always
35	Our analytics personnel work closely with customers and maintain productive user/client relationships	Strongly Disagree	Disagree	Moderately Disagree	Indifferent	Moderately Agree	Agree	Strongly Agree
36	We enforce adequate plans for the introduction and utilization of big data analytics	Never	Very Rarely	Rarely	Occasionally	Frequently	Very Frequently	Always
37	We perform big data analytics planning processes in systematic and formalized ways	Strongly Disagree	Disagree	Moderately Disagree	Indifferent	Moderately Agree	Agree	Strongly Agree
38	We frequently adjust big data analytics plans to better adapt to changing conditions	Strongly Disagree	Disagree	Moderately Disagree	Indifferent	Moderately Agree	Agree	Strongly Agree
39	When we make big data analytics investment decisions, we think about and estimate the effect they will have on the productivity of the employees' work	Never	Very Rarely	Rarely	Occasionally	Frequently	Very Frequently	Always
40	When we make big data analytics investment decisions, we think about and estimate the cost of training that end-users will need	Never	Very Rarely	Rarely	Occasionally	Frequently	Very Frequently	Always
41	In our organisation, business analysts and line people meet regularly to discuss important issues	Never	Very Rarely	Rarely	Occasionally	Frequently	Very Frequently	Always
42	In our organisation, business analysts and line people from various departments regularly attend cross-functional meetings	Never	Very Rarely	Rarely	Occasionally	Frequently	Very Frequently	Always

43	In our organisation, information is widely shared between business analysts and line people so that those who make decisions or perform jobs have access to all available know-how	Never	Very Rarely	Rarely	Occasionally	Frequently	Very Frequently	Always
44	In our organisation, the responsibility for big data analytics development is clear	Strongly Disagree	Disagree	Moderately Disagree	Indifferent	Moderately Agree	Agree	Strongly Agree
45	We are confident that big data analytics project proposals are properly appraised	Strongly Disagree	Disagree	Moderately Disagree	Indifferent	Moderately Agree	Agree	Strongly Agree
46	Our analytics department is clear about its performance criteria	Strongly Disagree	Disagree	Moderately Disagree	Indifferent	Moderately Agree	Agree	Strongly Agree
47	Our company is better than competitors in connecting (e.g., communication and information sharing) parties within a business process	Strongly Disagree	Disagree	Moderately Disagree	Indifferent	Moderately Agree	Agree	Strongly Agree
48	Our company is better than competitors in bringing detailed information into a business process	Never	Very Rarely	Rarely	Occasionally	Frequently	Very Frequently	Always
Section 6: Big data analytics Infrastructure Flexibility								
49	Compared to rivals within our industry, our organisation has the foremost available analytics systems	Strongly Disagree	Disagree	Moderately Disagree	Indifferent	Moderately Agree	Agree	Strongly Agree
50	All other (e.g., remote, branch, and mobile) offices are connected to the central office for analytics	Strongly Disagree	Disagree	Moderately Disagree	Indifferent	Moderately Agree	Agree	Strongly Agree
51	There are no identifiable communications bottlenecks within our organisation when sharing analytics insights	Strongly Disagree	Disagree	Moderately Disagree	Indifferent	Moderately Agree	Agree	Strongly Agree
52	Software applications can be easily transported and used across multiple analytics platforms	Never	Very Rarely	Rarely	Occasionally	Frequently	Very Frequently	Always
53	Our user interfaces provide transparent access to all platforms and applications	Never	Very Rarely	Rarely	Occasionally	Frequently	Very Frequently	Always
54	Reusable software modules are widely used in new analytics model development	Never	Very Rarely	Rarely	Occasionally	Frequently	Very Frequently	Always
55	The legacy system within our organisation restricts the development of new applications	Strongly Disagree	Disagree	Moderately Disagree	Indifferent	Moderately Agree	Agree	Strongly Agree
56	End-users can use software provided to create their own analytics applications	Never	Very Rarely	Rarely	Occasionally	Frequently	Very Frequently	Always
Section 7: Big data analytics Personnel Expertise Capability								
57	Our analytics personnel are very capable in terms of programming skills (e.g., structured programming, web-based application, CASE, tools, etc.)	Strongly Disagree	Disagree	Moderately Disagree	Indifferent	Moderately Agree	Agree	Strongly Agree

58	Our analytics personnel are very capable in terms of managing project life cycles	Strongly Disagree	Disagree	Moderately Disagree	Indifferent	Moderately Agree	Agree	Strongly Agree
59	Our analytics personnel are very capable in the areas of data and network management and maintenance	Strongly Disagree	Disagree	Moderately Disagree	Indifferent	Moderately Agree	Agree	Strongly Agree
60	Our analytics personnel are very capable in data decision support systems (e.g., expert systems, artificial intelligence, warehousing, mining, marts, etc.)	Strongly Disagree	Disagree	Moderately Disagree	Indifferent	Moderately Agree	Agree	Strongly Agree
61	Our analytics personnel are very knowledgeable about the critical factors for the success of our organisation	Strongly Disagree	Disagree	Moderately Disagree	Indifferent	Moderately Agree	Agree	Strongly Agree
62	Our analytics personnel are very capable in interpreting business problems and developing appropriate technical solutions	Strongly Disagree	Disagree	Moderately Disagree	Indifferent	Moderately Agree	Agree	Strongly Agree
63	Our analytics personnel are very knowledgeable about our business environment	Strongly Disagree	Disagree	Moderately Disagree	Indifferent	Moderately Agree	Agree	Strongly Agree
64	Our analytics personnel are very capable in terms of planning and executing work in a collective environment	Strongly Disagree	Disagree	Moderately Disagree	Indifferent	Moderately Agree	Agree	Strongly Agree
65	Our analytics personnel are very capable in terms of teaching others in our business	Strongly Disagree	Disagree	Moderately Disagree	Indifferent	Moderately Agree	Agree	Strongly Agree
Section 8: Firm Financial and Market Performance								
66	Using big data analytics improved customer retention during the last 3 years relative to competitors	Strongly Disagree	Disagree	Moderately Disagree	Indifferent	Moderately Agree	Agree	Strongly Agree
67	Using big data analytics improved Sales Growth during the last 3 years relative to competitors	Strongly Disagree	Disagree	Moderately Disagree	Indifferent	Moderately Agree	Agree	Strongly Agree
68	Using big data analytics improved Profitability during the last 3 years relative to competitors	Strongly Disagree	Disagree	Moderately Disagree	Indifferent	Moderately Agree	Agree	Strongly Agree
69	Using big data analytics improved Return on Investment (ROI) during the last 3 years relative to competitors	Strongly Disagree	Disagree	Moderately Disagree	Indifferent	Moderately Agree	Agree	Strongly Agree
70	Using big data analytics improved overall financial performance during the last 3 years relative to competitors	Strongly Disagree	Disagree	Moderately Disagree	Indifferent	Moderately Agree	Agree	Strongly Agree
71	Our success rate of new products or services has been higher than our competitors	Strongly Disagree	Disagree	Moderately Disagree	Indifferent	Moderately Agree	Agree	Strongly Agree
72	Using analytics our market share has exceeded that of our competitors	Strongly Disagree	Disagree	Moderately Disagree	Indifferent	Moderately Agree	Agree	Strongly Agree

9.1.3. Additional survey distribution data

Table 22: Forums used for Survey Distribution

Name	Population Size	Action taken	Access
LinkedIn Groups			
Big Data	2,041 members	Posted link and request to take survey on group page	https://www.linkedin.com/groups/8354642
Big Data and Analytics	351,094 members	Posted link and request to take survey on group page, Page owner endorsed link	https://www.linkedin.com/groups/4332669
Big Data, Analytics, Business Intelligence & Visualisation Experts Community	244,716 members	Posted link and request to take survey on group page	https://www.linkedin.com/groups/23006
<i>Harvard Business Review</i>	1,457,703 Members	Posted link and request to take survey on group page	https://www.linkedin.com/groups/3044917
MIT Big Data and Analytics (Official)	543 Members	Posted link and request to take survey on group page, Page owner endorsed link	https://www.linkedin.com/groups/8553978
Events, Conferences and Seminars			
Free Beer Session #23		Supervisor Advertised Link	http://marketingspread.co.za/2017/08/22/obsidian-systems-goes-wide-open-with-free-beer-session-no-23/
Predictive Analytics Masterclass		Supervisor Advertised Link	https://www.gibs.co.za/news-events/events/open-programmes/pages/predictive-analytics-masterclass.aspx
Executive ML (Machine Learning) conference and workshop		Attended conference and asked respondents to complete	https://www.executiveml.com/
Deep Learning Indaba		Attended conference and asked respondents to complete	http://www.deeplearningindaba.com/

9.1.4. Challenges of data collection by online survey

As described by Evans & Mathur (2005), though significant improvements in online surveys have been made in recent years, many concerns remain. These include declining response rates, incomplete or limited survey completion, as well as mistrust about application of surveys (actual research versus research for sale of data), survey length, privacy and security issues, blocking of access to surveys, perception of surveys being spam, and excessive interviewing by market resulting in respondent fatigue. Many of these issues appear to have influenced response rate within this study. In accordance with this, response rates were found to be initially limited. Due to this, the survey questionnaire was altered to emphasise that not all

questions were compulsory in order to further ease the completion of questions to which respondents did know the answers, and not simply cease the entire survey.

A contextually relevant concern in this study is the susceptibility of online surveys to non-response error (Evans & Mathur, 2005; Zikmund *et al.*, 2012). In the context of this study, this concerns the respondents' ability and/or individual perspective regarding the implementation and effect of BDA on FPer. These are factors that are relative to experience, position in business and exposure to the broader environment. As a result, respondents may not be able to provide informed responses to technically related or business related questions based on their own understanding or exposure. Response bias was a further potential challenge, where respondents had the potential to alter their answers in light of the environment and their participation of it, whether or it was a true reflection on their reality (Evans & Mathur, 2005).

Furthermore, the completeness of online surveys cannot be guaranteed in online surveys, as the researcher would not be present to enforce it. Given the nature of the format, a lower level of detail is furthermore typically collected, which may result in nuance being lost in some cases (Zikmund *et al.*, 2012). These concerns were mitigated to an extent by the researcher applying as detailed and personalised a communication as was possible with respondents prior to them receiving the survey link, including telephonic and personalised email communication to engender good will for improved response (Zikmund *et al.*, 2012). Response bias was further mitigated by ensuring complete anonymity of respondents, which was anticipated to support more thorough and truthful responses, despite it mitigating the ability to follow up and request given surveys be completed.

9.2. Appendix 2: Ethical Clearance

**Gordon
Institute
of Business
Science**
University
of Pretoria

20 July 2017

Michael Niland

Dear Michael,

Please be advised that your application for Ethical Clearance has been approved.

You are therefore allowed to continue collecting your data.

We wish you everything of the best for the rest of the project.

Kind Regards

GIBS MBA Research Ethical Clearance Committee

9.3. Appendix 3: Data Preparation

9.3.1. Data Collected

Data was exported from SurveyMonkey™ as an Excel document with English language responses provided per respondent. A sample extract is provided below (Table 23) illustrating responses to Section 1 of the questionnaire (Table 21).

Table 23: Sample extract of data collected using SurveyMonkey (only a portion of respondents and only data up to Section 1 are illustrated)

Respondent Number	Respondent ID	Collector ID	Start Date	End Date	Are you aware of or associated with a Big Data Analytics Capability within the organisation being described in this questionnaire?	Which of the following best describes the principal industry of your organisation?	How long have you been associated with the field of big data or data analytics?	How long has your organisation actively pursued or applied big data analytics to its business?	What is your main association with the data analytics capability?	What is the approximate total number of employees within your organisation?	Which of the following best describes your current job level?	In what country do you work?
1	637090432	160102819	2017-09-04 15:40:32	2017-09-01 11:43:51	Yes	Food & Beverages	0-2	0+	Data analyst (Direct process)	1,000 or more	Intermediate	South Africa
2	636941932	160102819	2017-08-31 22:45:37	2017-08-31 22:54:25	Yes	Finance and Financial Services	15+	4-6	User of analytics within business	1,000 or more	Middle Management	South Africa
3	636941932	160102819	2017-08-31 22:45:37	2017-08-31 22:54:25	Yes	Finance and Financial Services	5-10	4-6	User of analytics within business	1,000 or more	Middle Management	South Africa
4	636942479	160102819	2017-08-31 22:53:45	2017-08-31 22:59:08	Yes	Finance and Financial Services	0-2	0-2	Data analyst (Direct process)	1,000 or more	Senior Management	South Africa
5	636946694	160102819	2017-08-31 10:42:49	2017-08-31 10:57:59	Yes	Finance and Financial Services	5-10	0+	User of analytics within business	1,000 or more	Middle Management	South Africa
6	636930632	160102819	2017-08-30 14:52:13	2017-08-30 15:08:25	Yes	Transportation & Delivery	0-2	0-2	Data analyst (Direct process)	1,000 or more	Middle Management	South Africa
7	6366151376	160102819	2017-08-30 13:05:03	2017-08-30 13:16:34	Yes	Telecommunications, Techno	0-2	0-2	Big Data Management (Driver)	500-999	Senior Management	South Africa
8	636920251	160102819	2017-08-30 20:30:32	2017-08-30 20:42:45	Yes	Telecommunications, Techno	5-10	0-2	Data analyst (Direct process)	1,000 or more	Senior Management	South Africa
9	636923344	160102819	2017-08-29 23:20:14	2017-08-29 23:35:21	Yes	Telecommunications, Techno	2-5	0+	IT Systems or infrastructure	1,000 or more	Owner/Executive/C-Level	South Africa
10	6362054698	160102819	2017-08-28 16:23:39	2017-08-28 16:35:34	Yes	Telecommunications, Techno	0-2	4-6	Advising customers about data	1,000 or more	Intermediate	Luxembourg
11	636179510	160102819	2017-08-28 13:20:34	2017-08-28 13:31:01	Yes	Finance and Financial Services	2-5	0-2	User of analytics within business	500-999	Owner/Executive/C-Level	United Arab Emirates
12	636179510	160102819	2017-08-28 13:20:34	2017-08-28 13:31:01	Yes	Insurance	12-5	4-6	IT Systems or infrastructure	1,000 or more	Senior Management	Ireland
13	636097743	160102819	2017-08-27 7:36:23	2017-08-27 7:49:29	Yes	Manufacturing	0-2	0+	User of analytics within business	1,000 or more	Middle Management	South Africa
14	6358669084	160102819	2017-08-25 15:40:31	2017-08-25 15:51:52	Yes	Insurance	2-5	6-8	Data analyst (Direct process)	500-999	Intermediate	South Africa
15	635862714	160102819	2017-08-25 14:14:38	2017-08-25 15:52:08	Yes	Insurance	0-2	0-2	Data analyst (Direct process)	500-999	Intermediate	South Africa
16	6358497067	160102819	2017-08-25 13:49:14	2017-08-25 14:08:12	Yes	Telecommunications, Techno	2-5	0-2	IT Systems or infrastructure	500-999	Middle Management	South Africa
17	6358204462	160102819	2017-08-25 8:27:14	2017-08-25 8:37:07	Yes	Finance and Financial Services	5-10	4-6	User of analytics within business	500-999	Intermediate	South Africa
18	635629318	160102819	2017-08-24 14:43:26	2017-08-24 14:52:52	Yes	Telecommunications, Techno	0-2	0-2	User of analytics within business	1-99	Middle Management	South Africa
19	635648817	160102819	2017-08-24 14:13:45	2017-08-24 14:29:38	Yes	Retail & Consumer Durables	0-2	0-2	User of analytics within business	1-99	Owner/Executive/C-Level	South Africa
20	6355446343	160102819	2017-08-23 21:32:28	2017-08-23 21:47:43	Yes	Finance and Financial Services	5-10	0+	IT Systems or infrastructure	1,000 or more	Senior Management	South Africa
21	6355110946	160102819	2017-08-23 18:46:04	2017-08-23 18:57:02	Yes	Utilities, Energy, And Extract	2-5	6-8	IT Systems or infrastructure	1,000 or more	Senior Management	South Africa
22	6347401349	160102819	2017-08-23 12:55:58	2017-08-23 13:41:25	Yes	Telecommunications, Techno	0-2	0-2	Big Data Management (Driver)	500-999	Owner/Executive/C-Level	South Africa
23	634723217	160102819	2017-08-18 7:54:18	2017-08-18 8:20:18	Yes	Agriculture	15-10	4-6	IT Systems or infrastructure	1,000 or more	Owner/Executive/C-Level	South Africa
24	634638627	160102819	2017-08-16 12:37:16	2017-08-17 16:11:23	Yes	Finance and Financial Services	5-10	4-6	Data analyst (Direct process)	500-999	Intermediate	South Africa
25	6345126942	160102819	2017-08-17 6:14:44	2017-08-17 6:14:34	Yes	Telecommunications, Techno	0-2	0-2	IT Systems or infrastructure	500-999	Senior Management	South Africa
26	6343166578	160102819	2017-08-16 03:55	2017-08-16 8:17:03	Yes	Telecommunications, Techno	2-5	2-4	IT Systems or infrastructure	1,000 or more	Middle Management	South Africa
27	6343148305	160102819	2017-08-16 7:41:57	2017-08-16 8:09:23	Yes	Finance and Financial Services	2-5	0+	Data analyst (Direct process)	1,000 or more	Intermediate	South Africa
28	633995571	160102819	2017-08-11 17:21:09	2017-08-11 17:33:05	Yes	Telecommunications, Techno	2-5	0-2	Big Data Management (Driver)	1,000 or more	Owner/Executive/C-Level	South Africa
29	6338484056	160102819	2017-08-11 15:42:22	2017-08-11 16:20:34	Yes	Telecommunications, Techno	0-2	0-2	Data analyst (Direct process)	Don't know	Entry Level	South Africa
30	633813448	160102819	2017-08-11 15:41:12	2017-08-11 15:49:23	Yes	Telecommunications, Techno	2-5	0-2	Data analyst (Direct process)	1,000 or more	Senior Management	South Africa
31	6336969723	160102819	2017-08-11 14:17:42	2017-08-11 14:30:50	Yes	Telecommunications, Techno	5-10	0-2	Privacy and Security	1,000 or more	Senior Management	South Africa
32	633639833	160102819	2017-08-11 13:30:13	2017-08-11 13:42:30	Yes	Telecommunications, Techno	0-2	0-2	User of analytics within business	500-999	Middle Management	South Africa

9.3.2. Data Coding

In order to facilitate assessment of data by statistical programmes, English language responses provided by respondents needed to be converted to numeric values. This was initially done according to Table 24 below, wherein single numeric values were associated with each possible answer option given in the data collected.

An illustration of the result of coding is provided in Table 25 below, wherein responses of section 2 and section 3 of the questionnaire (Table 21) were coded. Section 1 was not coded as it allowed for descriptive characteristics of respondents to be applied in imputation of data. The responses which required imputation are highlighted cells in Table 25 below.

Table 24: Coding of English language variables to numeric values

Strongly Disagree	1	Yes	22	2 – 4	44	Entry Level	63
Disagree	2	No	23	4 – 6	45	Intermediate	64
Moderately Disagree	3	Food & Beverages	24	8+	46	Middle Management	65
Indifferent	4	Finance and Financial Services	25	6 – 8	47	Senior Management	66
Moderately Agree	5	Transportation & Delivery	26	User of analytics within business	48	Owner/Executive/C-Level	67
Moderately Agree	5	Telecommunications, Technology, Internet & Electronics	27	Data analyst (Direct processor of data)	49	South Africa	68
Agree	6	Insurance	28	Big Data Management (Driving application of resources)	50	Luxembourg	69
Strongly Agree	7	Manufacturing	29	IT Systems or Infrastructure (Data technology environment)	51	United Arab Emirates	70
Never	8	Insurance	30	Advising customers about data analytics capability	52	Ireland	71
Very Rarely	9	Retail & Consumer Durables	31	Privacy and Security	53	United Kingdom of Great Britain and Northern Ireland	72
Rarely	10	Utilities, Energy, And Extraction	32	end to end	54	Belgium	73
Occasionally	11	Agriculture	33	Manage development teams - maybe Big Data Management	55		
Frequently	12	Automotive	34	CTO	56		
Very Frequently	13	Nonprofit	35	Consultant	57		
Always	14	Education	36	1 – 99	58		
None of them	15	Business Support & Logistics	37	100 – 499	59		
2	16	I am currently not employed	38	500 – 999	60		
3	17	0 – 2	39	1,000 or more	61		
Some of them	18	2 – 5	40	Don't know	62		
5	19	5 – 10	41				
6	20	10 – 15	42				
All of them	21	15+	43				
Yes	22						
No	23						

Table 25: Illustrative results of coding processes

[illegible]

9.3.3. Data Imputation

As informed by King *et al.* (2001), Schafer & Graham (2002) and Scheffer (2002), missing data in surveys with at least 50% of questions completed were filled in or completed through imputation using averaged responses. The summary of the data collected is illustrated through Table 26, which describes both the raw level of responses, and the extent of imputation across the data set.

More specifically, Imputation was done by averaging responses per question from a group of respondents which portrayed similar characteristics to the respondent with a missing question. Importantly, the relevant group was determined through filtering for descriptive traits of that respondent (Section 1 of the questionnaire in Appendix 1, Table 21) that best related to the question being asked. This is best clarified by way of example; if the question related to competitors then groups which were then used to impute a given missing answer were determined by filtering by the respondents who was missing data's industry (such as telecommunication or insurance respectively). Similarly, if a given question with missing data considered technical details pertaining to a respondents business, groups were filtered by type of role that that respondent had (managerial, executive, analyst, systems engineer or user of data). This then allowed for the most appropriate grouping from which responses could be averaged to be acquired from correctly completed data that was available in the data set.

Table 26: Summary results of data collected and data imputed

Number of questions with missing answers per respondent	Number of respondents per number of questions with missing answers	% of Questions completed per respondent	% of Total Respondents	% of Respondents (80) used in final analysis	Number of Answers Imputed
0	51	100.00%	46.79%	63.75%	0
1	6	98.61%	5.50%	7.50%	6
2	4	97.22%	3.67%	5.00%	8
3	3	95.83%	2.75%	3.75%	9
4	2	94.44%	1.83%	2.50%	8
5	1	93.06%	0.92%	1.25%	5
6	1	91.67%	0.92%	1.25%	6
7	2	90.28%	1.83%	2.50%	14
8	1	88.89%	0.92%	1.25%	8
10	3	86.11%	2.75%	3.75%	30
14	1	80.56%	0.92%	1.25%	14
18	1	75.00%	0.92%	1.25%	18
20	1	72.22%	0.92%	1.25%	20
30	1	58.33%	0.92%	1.25%	30
32	1	55.56%	0.92%	1.25%	32
35	1	51.39%	0.92%	1.25%	35
48	1	33.33%	0.92%		
50	1	30.56%	0.92%		
59	1	18.06%	0.92%		
60	0	16.67%	0.00%		
61	0	15.28%	0.00%		
62	0	13.89%	0.00%		
63	0	12.50%	0.00%		
64	24	11.11%	22.02%		

This was carried out where respondent 79 did not answer question 10, (cell highlighted in Table 27). In order to complete the result, relevant characteristics pertaining to the question being asked were selected. As question 10 asked “Our organisation enables learning, accumulation and application of new knowledge better than our competitors”, it related to an industry. Thus the industry of that respondent was selected as a grouping term (Insurance in this case), and the average of the results reported (5.55) in this case. As whole numbers were

required, this was rounded to 6, which corresponds to “Agree” as an imputed result. This process was repeated through the entire table to ‘complete’ questionnaires.

Table 27: Selection of grouping for imputation of a missing answer in the highlighted cell

	Are you aware of or associated with a Big Data Analytics Capability within the organisation being described in this questionnaire?	Which of the following best describes the principal industry of your organisation?	How long have you been associated with the field of big data or data analytics?	How long has your organisation actively pursued or applied big data analytics to its business?	What is your main association with the data analytics capability?	What is the approximate total number of employees within your organisation?	Which of the following best describes your current job level?	In what country do you work?	Our organisation has a widely held belief that innovation is an absolute necessity for the organisation's future	Our organisation enables learning, accumulation and application of new knowledge better than our competitors	We believe it is important to adopt new and cutting-edge practices to continuously improve product or service delivery	People in our organisation are continuously encouraged to expand their capacities to achieve more and apply new capabilities
Respondent Number												
12	Yes	Insurance	2 – 5	4 – 6	IT Systems or Infrastructure	1,000 or more	Senior Management	Ireland	7	5	6	12
14	Yes	Insurance	2 – 5	6 – 8	Data analyst (Direct pre)	500 – 999	Intermediate	South Africa	6	6	6	13
15	Yes	Insurance	0 – 2	0 – 2	Data analyst (Direct pre)	100 – 499	Intermediate	South Africa	2	3	5	11
41	Yes	Insurance	5 – 10	0 – 2	Data analyst (Direct pre)	100 – 499	Middle Management	South Africa	5	6	3	11
58	Yes	Insurance	2 – 5	2 – 4	User of analytics within	500 – 999	Senior Management	South Africa	7	6	6	11
61	Yes	Insurance	0 – 2	0 – 2	User of analytics within	100 – 499	Intermediate	South Africa	7	7	7	14
66	Yes	Insurance	2 – 5	0 – 2	User of analytics within	1,000 or more	Intermediate	South Africa	7	6	7	13
77	Yes	Insurance	0 – 2	0 – 2	User of analytics within	1,000 or more	Intermediate	South Africa	6	5.555555556	6	12
81	Yes	Insurance	2 – 5	0 – 2	Data analyst (Direct pre)	1,000 or more	Middle Management	South Africa	6	6	5	12
83	Yes	Insurance	5 – 10	0 – 2	IT Systems or Infrastructure	1,000 or more	Middle Management	South Africa	6	5	6	14
87	No	Insurance	5 – 10	4 – 6	IT Systems or Infrastructure	1,000 or more	Middle Management	United Kingdom of Great Britain	0	0	0	0
108	No	Insurance	2 – 5	8+	Data analyst (Direct pre)	100 – 499	Middle Management	South Africa	0	0	0	0

This process was replicated across the entire data set for each question where and answer was missing, as long as the total number of responses per respondent was higher than 50%. The summary of these actions are described through Table 3.

The complete table of coded and imputed (only the top 80 respondents who answered affirmatively to the screening question and who had completed answers to more than 50% of questions) is provided as Table 28.

Table 28: Complete Coded and Imputed Data Set

Respondent Number	Are you aware of or associated with a Big Data Analytics Capability within the organisation being described in this questionnaire?	Which of the following best describes the principal industry of your organisation?	How long have you been associated with the field of big data analytics?	How long has your organisation actively pursued or applied its business?	What is your main association with the data analytics capability?	What is the approximate total number of employees within your organisation?	Which of the following best describes your current job level?	In what country do you work?	Our organisation has a 'today/told' belief that innovation is an absolute necessity for the organisation's future	Our organisation practices innovation, accumulation of new knowledge better than our competitors	We believe it is important to adopt new and cutting-edge practices to continuously improve product or service delivery	People in our organisation are encouraged to expand their capacities to achieve more and apply new capabilities	Our organisation can be described as visionary and flexible	There is an extensive employee orientation program for new employees to ensure corporate vision and purpose	We invest in targeted training and support at all levels of our organisation to assist our organisation to understand or know how to use data that is available	Our executive level support our big data analytics capability	Before any decision is taken we systematically evaluate internal data to better understand the nature of the problem and what to do about it	Employees receive recognition from the organisation for applying evidence based decisions making in our typical business processes	Managers in our organisation tend to believe that experience and knowledge gained on the job is the only important source of information when considering how to tackle a problem	We make decisions by looking at what other organisations are doing, and how they are working for them	Internal politics and power struggles influence the way we make decisions about policies and practices	We believe that our competitiveness depends on our analytics capability	Our decision makers have access to a management information system	Our managers know how to critically appraise both internal data and evidence from scientific research	
1	22	24	39	46	49	61	64	68	7	4	1	1	1	4	4	4	11	10	4	9	13	6	12	18	
2	22	25	39	46	49	61	64	68	7	4	1	1	1	4	4	11	10	4	9	13	6	12	18		
3	22	25	41	45	48	61	65	68	7	6	1	1	1	4	4	11	13	4	9	13	5	13	19		
4	22	25	41	45	48	61	65	68	7	6	1	1	1	4	4	11	13	4	9	13	5	13	19		
5	22	25	41	45	48	61	65	68	7	6	1	1	1	4	4	11	13	4	9	13	5	13	19		
6	22	26	39	46	49	61	66	68	5	4	5	11	2	1	5	3	8	9	4	11	14	2	14	20	
7	22	27	39	46	49	61	66	68	5	4	5	11	2	1	5	3	8	9	4	11	14	2	14	20	
8	22	27	41	45	48	61	66	68	5	4	5	11	2	1	5	3	8	9	4	11	14	2	14	20	
9	22	27	39	46	49	61	66	68	5	4	5	11	2	1	5	3	8	9	4	11	14	2	14	20	
10	22	27	39	46	49	61	66	68	5	4	5	11	2	1	5	3	8	9	4	11	14	2	14	20	
11	22	25	40	39	48	59	67	70	7	6	1	1	1	4	4	11	14	3	8	14	6	14	21		
12	22	28	40	45	51	61	66	71	7	5	1	1	1	4	4	11	14	3	8	14	6	14	21		
13	22	29	39	46	48	61	65	68	5	3	5	11	1	3	6	12	13	3	11	12	2	14	18		
14	22	28	40	47	49	60	64	68	6	6	6	13	6	7	6	13	6	13	6	13	6	12	20		
15	22	28	39	39	49	59	64	68	2	3	3	11	1	2	2	3	13	10	5	10	13	2	9	18	
16	22	27	40	39	51	60	64	68	6	6	6	13	6	4	4	4	11	11	5	11	12	6	12	18	
17	22	25	41	45	48	60	64	68	5	1	1	6	9	2	1	2	2	10	8	7	10	14	3	13	16
18	22	27	39	39	48	58	65	68	5	5	5	12	5	6	4	5	11	11	6	11	13	4	10	16	
19	22	31	39	48	58	67	68	6	7	5	7	12	7	6	5	7	13	12	3	10	9	6	12	18	
20	22	32	41	46	51	61	66	68	7	6	7	14	5	7	7	13	12	2	9	10	6	12	18		
21	22	32	41	46	51	61	66	68	7	6	7	14	5	7	7	13	12	2	9	10	6	12	18		
22	22	32	41	46	51	61	66	68	7	6	7	14	5	7	7	13	12	2	9	10	6	12	18		
23	22	33	41	45	51	61	67	68	3	5	5	13	6	6	2	5	13	12	2	11	8	4	14	18	
24	22	33	41	45	51	61	67	68	3	5	5	13	6	6	2	5	13	12	2	11	8	4	14	18	
25	22	27	39	39	48	58	65	68	7	6	6	14	5	7	7	13	12	2	9	10	6	12	18		
26	22	27	39	39	48	58	65	68	7	6	6	14	5	7	7	13	12	2	9	10	6	12	18		
27	22	21	40	39	50	61	64	68	6	6	6	14	5	7	7	13	12	2	9	10	6	12	18		
28	22	27	39	39	50	61	64	68	7	5	5	12	6	2	7	11	11	3	12	11	3	9	16		
29	22	27	39	39	49	59	67	68	6	6	6	13	6	2	7	11	11	3	12	11	3	9	16		
30	22	27	40	39	49	61	66	68	6	5	5	13	7	7	5	6	13	13	3	14	12	7	10	15	
31	22	27	39	39	48	60	65	68	6	5	5	13	7	7	5	6	13	13	3	14	12	7	10	15	
32	22	27	39	39	48	60	65	68	7	3	7	13	5	1	2	7	9	11	6	11	14	4	13	16	
33	22	25	42	39	48	60	66	68	5	3	3	10	3	2	2	6	11	10	4	11	12	2	9	16	
34	22	25	42	39	48	60	66	68	5	3	3	10	3	2	2	6	11	10	4	11	12	2	9	16	
35	22	27	42	47	58	67	68	7	4	5	5	13	4	4	5	7	14	14	5	12	11	6	13	20	
36	22	31	41	39	48	59	67	68	4	5	5	13	4	4	5	7	14	14	5	12	11	6	13	20	
37	22	31	41	39	54	58	67	68	1	2	7	14	7	4	5	7	14	14	3	11	11	5	10	16	
38	22	26	39	39	50	61	65	68	6	2	6	14	5	1	5	9	8	3	11	14	4	12	18		
39	22	25	40	45	48	61	65	68	7	7	7	14	7	7	7	14	14	1	9	9	7	14	18		
40	22	33	40	45	48	61	65	68	7	7	7	14	7	7	7	14	14	1	9	9	7	14	18		
41	22	28	39	39	49	59	65	68	3	3	3	11	3	4	5	6	12	3	11	12	6	11	17	20	
42	22	25	40	45	48	61	66	68	7	7	7	14	7	7	7	14	14	1	9	9	7	14	18		
43	22	27	39	39	50	61	66	68	7	6	6	14	5	7	7	13	12	2	10	13	6	13	20		
44	22	27	39	39	50	61	66	68	7	6	6	14	5	7	7	13	12	2	10	13	6	13	20		
45	22	27	39	39	50	61	66	68	7	6	6	14	5	7	7	13	12	2	10	13	6	13	20		
46	22	27	39	39	50	61	66	68	7	6	6	14	5	7	7	13	12	2	10	13	6	13	20		
47	22	27	39	39	50	61	66	68	7	6	6	14	5	7	7	13	12	2	10	13	6	13	20		
48	22	31	40	47	48	61	64	68	3	5	5	12	2	2	3	12	11	4	12	2	13	6	13	19	
49	22	31	40	47	48	61	64	68	3	5	5	12	2	2	3	12	11	4	12	2	13	6	13	19	
50	22	27	39	39	50	60	66	68	7	5	5	13	3	4	3	5	12	12	4	12	11	4	12	18	
51	22	27	39	39	50	60	66	68	7	5	5	13	3	4	3	5	12	12	4	12	11	4	12	18	
52	22	27	39	39	50	60	66	68	7	5	5	13	3	4	3	5	12	12	4	12	11	4	12	18	
53	22	26	39	39	44	51	61	64	6	4	6	13	6	2	1	2	11	13	4	9	11	7	12	17	
54	22	26	40	39	48	61	65	68	7	6	7	13	6	2	1	2	11	13	4	9	11	7	12	17	
55	22	25	40	39	56	58	67	68	7	5	7	13	6	3	1	6	12	9	10	6	11	5	10	16	
56	22	27	40	39	56	58	67	68	7	5	7	13	6	3	1	6	12	9	10	6	11	5	10	16	
57	22	27	39	39	48	58	67	68	6	5	5	13	5	6	5	6	12	11	5	13	12	5	13	18	
58	22	28	44	44	48	60	66	68	7	6	6	11	5	5	6	6	13	11	6	11	11	6	12	19	
59	22	34	39	39	48	58	67	68	4	5	5	11	3	5	11	3	11	3	11	13	4	11	16	14	
60	22	35	41	45	48	61	66	68	7	6	6	14	5	7	7	13	12	3	10	13	6	13	20		
61	22	28	39	39	48	59	64	68	7	7	7	14	7	7	6	4	13	13	4	12	11	6	14	21	
62	22	27	41	44	49	61	65	68	5	6	5	11	5	6	5	6	13	12	5	12	11	5	12	19	
63	22	27	41	44	49	61	65	68	6	6	6	11	3	4	4	11	11	4	10	13	4	11	18	14	
64	22	27	40	39	49	58	67	68	7	5	5	13	7	2	7	5	12	11	1	9	8	6	13	17	
65	22	25	41	46	48	58	66	68	6	5	5	13	5	6	5	6	13	13	5	12	13	7	11	20	
66	22	28	40	39	48	61	64	68	7	6	7	13	7	7	7	5	13	12	4	11	7	8	14	20	
67	22	27	39	39	51	60	64	68	6	6	6	13	5	5	4	6	11	11	5	11	5	10	16	14	
68	22	31	39	45	48	61	64	68	6	3	7	11	6	3	3	5	11	13	5	11	5	5	11	18	
69	22	25	40	46	48	61	65	68	5	4	5	12	5	4	4	11	11	6	12	13	4	11	16	14	
70	22	26	39	39	50	61	65	68	7	6	6	14	5	7	7	13	12	2	10	13					

152

153

9.4. Appendix 4: Data Analysis

9.4.1. Descriptive analyses of data

Table 29: Assessment of descriptive statistics for each question

Descriptive Statistics									
		N	Minimum	Maximum	Mean	Std. Deviation	Variance	Skewness	
Section 1: Context of Organisation and Respondent									
know_bda	Q1	80	1	1	1.00	0.000	0.000		
industry	Q2	80	1	13	4.35	2.576	6.635	1.446	0.269
time_associated_with_bda	Q3	80	1	5	2.08	1.053	1.108	0.849	0.269
time_org_in_bda	Q4	80	1	5	2.26	1.532	2.348	0.778	0.269
association_to_bda	Q5	80	1	8	2.39	1.471	2.164	1.352	0.269
number_employees	Q6	80	1	5	3.21	1.133	1.283	-0.968	0.269
current_job_level	Q7	80	1	5	3.38	1.118	1.250	-0.012	0.269
country	Q8	80	1	4	1.08	0.414	0.172	6.016	0.269
Section 2: Organisational Culture to Big Data Analytics Capability									
innovation_nb_org_future	Q9	80	1	7	6.10	1.176	1.382	-2.068	0.269
org_help_learning_app_knowledge	Q10	80	1	7	4.95	1.404	1.972	-0.697	0.269
adopt_new_practice_continuously	Q11	80	3	7	5.90	0.963	0.927	-0.756	0.269
org_encourage_staff_new_capabilities	Q12	80	2	7	5.14	1.319	1.740	-0.192	0.269
org_visionary_flexible	Q13	80	1	7	4.95	1.622	2.630	-0.721	0.269
employee_orientation_program	Q14	80	1	7	4.39	1.761	3.101	-0.415	0.269
org_trains_use_of_data	Q15	80	1	7	4.08	1.682	2.830	-0.203	0.269
exec_support_bda	Q16	80	1	7	5.34	1.458	2.125	-0.838	0.269
Section 3: Data Driven Decision Making in the Organisation									
understand_internal_data_to_clarify_problem	Q17	80	1	7	4.96	1.267	1.606	-0.464	0.269
org_recognise_staff_dddm	Q18	80	1	7	4.38	1.335	1.782	-0.036	0.269
org_believe_experience_n_b	Q19	80	1	7	4.03	1.359	1.847	-0.232	0.269
org_decision_based_on_other_orgs	Q20	80	2	7	4.10	1.098	1.205	-0.026	0.269
internal_politics_drive_org_decisions	Q21	80	1	7	4.63	1.594	2.541	-0.551	0.269
org_believe_competitiveness_from_analytics	Q22	80	1	7	4.84	1.579	2.492	-0.873	0.269
decision_makers_access_org_data	Q23	80	2	7	5.00	1.458	2.127	-0.276	0.269
managers_appraise_information	Q24	80	1	7	4.14	1.348	1.816	0.062	0.269
Section 4: Dynamic Capability / Adsorptive Capacity of the Organisation									
bureaucratic_org	Q25	80	1	7	4.59	1.784	3.182	-0.394	0.269
org_structure_flexible_and_allows_resource_sharing	Q26	80	1	7	4.43	1.644	2.703	-0.416	0.269

more_prod_services_than_comp	Q27	80	1	7	4.55	1.698	2.884	-0.502	0.269
quick_infusion_new_capabilities_in_org	Q28	80	1	7	3.95	1.882	3.542	-0.218	0.269
better_comms_than_comp	Q29	80	1	7	3.95	1.517	2.301	-0.114	0.269
Prod_service_change_fast	Q30	80	1	7	5.08	1.613	2.602	-0.626	0.269
org_dev_resources_better_comp	Q31	80	1	7	4.08	1.376	1.893	-0.407	0.269
Section 5: Big Data Analytics Management Capabilities									
continuous_assess_strat_appn_bda	Q32	80	2	7	5.01	1.196	1.430	-0.662	0.269
estimate_manager_time_overseeing_bda_investment	Q33	80	1	7	3.84	1.479	2.188	0.215	0.269
project_users_speed_up_decision_making	Q34	80	1	7	4.59	1.411	1.992	-0.418	0.269
bda_team_close_to_users	Q35	80	1	7	4.79	1.473	2.169	-0.841	0.269
plan_for_intro_of_bda	Q36	80	1	7	4.14	1.329	1.766	-0.159	0.269
bda_planned_systematically	Q37	80	1	7	4.29	1.416	2.005	-0.665	0.269
bda_plan_adjusted_to_conditions	Q38	80	1	7	4.44	1.457	2.123	-0.830	0.269
estimate_bda_impact_on_productivity	Q39	80	1	7	4.25	1.488	2.215	-0.136	0.269
estimate_training_cost	Q40	80	1	7	4.08	1.421	2.020	0.000	0.269
bda_meet_line_people_regularly	Q41	80	2	7	4.56	1.330	1.768	-0.138	0.269
cross_functional_meetings_norm	Q42	80	2	7	4.24	1.265	1.601	0.076	0.269
infor_shared_from_bda_to_line	Q43	80	1	7	4.33	1.199	1.437	-0.388	0.269
bda_dev_responsibility_clear	Q44	80	1	7	4.31	1.539	2.369	-0.245	0.269
bda_proposal_properly_appraised	Q45	80	1	7	4.35	1.332	1.775	-0.179	0.269
bda_dept_clear_performance_criteria	Q46	80	1	7	4.56	1.457	2.123	-0.530	0.269
org_link_people_to_business_process_well	Q47	80	1	7	4.40	1.365	1.863	-0.826	0.269
org_convert_info_to_business_process	Q48	80	1	7	4.38	1.247	1.554	-0.431	0.269
Section 6: Big data analytics Infrastructure Flexibility									
better_bda_than_comp	Q49	80	1	7	4.39	1.488	2.215	-0.532	0.269
remote_branch_connected_bda	Q50	80	1	7	4.59	1.697	2.878	-0.425	0.269
bottleneck_to_sharing_bda_insight	Q51	80	1	7	3.75	1.761	3.101	0.036	0.269
software_transplanted	Q52	80	1	7	3.94	1.453	2.110	0.162	0.269
users_access_all	Q53	80	1	7	4.06	1.461	2.135	0.014	0.269
software_reusable	Q54	80	1	7	4.30	1.570	2.466	-0.595	0.269
legacy_system_restricts	Q55	80	1	7	4.68	1.756	3.083	-0.479	0.269
users_do_own_analytics	Q56	80	1	7	3.64	1.678	2.816	-0.113	0.269

Section 7: Big data analytics Personnel Expertise Capability									
bda_good_programmers	Q57	80	1	7	5.60	1.308	1.711	-1.268	0.269
bda_mange_project_lifecycle	Q58	80	1	7	5.00	1.350	1.823	-0.633	0.269
bda_capable_data_and_network_maintenance	Q59	80	2	7	5.10	1.228	1.509	-0.573	0.269
bda_capable_data_decision_systems	Q60	80	2	7	5.10	1.289	1.661	-0.555	0.269
bda_know_business_critical_success_factors	Q61	80	2	7	5.28	1.125	1.265	-0.786	0.269
bda_make_good_solution_for_business_needs	Q62	80	2	7	5.28	1.242	1.544	-0.950	0.269
bda_understand_bussiness_environment	Q63	80	2	7	5.29	1.150	1.321	-0.897	0.269
bda_work_in_team	Q64	80	2	7	5.11	0.994	0.987	-0.707	0.269
bda_teach_well	Q65	80	1	7	4.86	1.473	2.171	-0.780	0.269
Section 8: Firm Financial and Market Performance									
bda_improved_customer_retention_vs_comp	Q66	80	1	7	4.58	1.412	1.994	-0.502	0.269
bda_improved_sales_vs_comp	Q67	80	1	7	4.60	1.298	1.686	-0.386	0.269
bda_improved_profitability_vs_comp	Q68	80	1	7	4.65	1.397	1.952	-0.432	0.269
bda_improved_ROI_vs_comp	Q69	80	1	7	4.44	1.320	1.743	-0.318	0.269
bda_improved_fin_performance_vs_comp	Q70	80	1	7	4.45	1.368	1.871	-0.258	0.269
bda_improved_new_product_success_vs_comp	Q71	80	1	6	4.26	1.319	1.740	-0.807	0.269
bda_improved_market_share_vs_comp	Q72	80	1	7	4.21	1.384	1.916	-0.246	0.269

9.4.2. Assessment of internal reliability (Cronbach's alpha)

The following table provides an illustrative data set of that which was carried out on all constructs and data applied in the study. As Cronbach's alpha analysis of data from Section 3 (DDDM) initially revealed an $\alpha = 0.449$, subsequent assessment of component questions was required to determine which should be removed to afford reliability of the construct.

Table 30: Example of Cronbach alpha item-total statistics to illustrate which items to delete to increase total reliability score

Item-Total Statistics					
Question Number	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
q17	31.10	19.408	0.391	0.512	0.336
q18	31.69	17.281	0.569	0.572	0.246
q19	32.04	25.328	-0.132	0.212	0.542
q20	31.96	20.897	0.326	0.192	0.373
q21	31.44	29.237	-0.371	0.371	0.653
q22	31.23	19.139	0.271	0.287	0.379
q23	31.06	18.743	0.357	0.280	0.339
q24	31.93	17.640	0.523	0.460	0.267

As illustrated in Table 31, the final column indicated α score if an item were to be deleted. According to this, the bold cell described the best possible result if an item were to be deleted. Question 21 was thus deleted from the analysis, and the assessment repeated until $\alpha > 0.7$.

Resulting α scores were derived with $N_{\text{respondents}} = 80$ in all cases, all with 100% cases valid. $N_{\text{sample size}}$ varied in each case, depending on the number of questions included in the construct. In each case mean and standard deviation were also calculated.

Table 31: Representative Cronbach's alpha assessment of Section 3 (DDDM) of the survey

Case Processing Summary			
		N	%
Cases	Valid	80	100.0
	Excluded ^a	0	0.0
	Total	80	100.0

Reliability Statistics		
Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
0.772	0.771	6

Item Statistics			
	Mean	Std. Deviation	N
q17_understand_internal_data_to_clarify_problem	4.96	1.267	80
q18_org_recognise_staff_dddm	4.38	1.335	80
q20_org_decision_based_on_other_orgs	4.10	1.098	80
q22_org_believe_competitiveness_from_analytics	4.84	1.579	80
q23_decision_makers_access_org_data	5.00	1.458	80
q24_managers_appraise_information	4.14	1.348	80

9.4.3. Assessment of construct internal validity by Pearson's correlation

Table 32: Confirmation of internal validity per construct through two Pearson's correlation

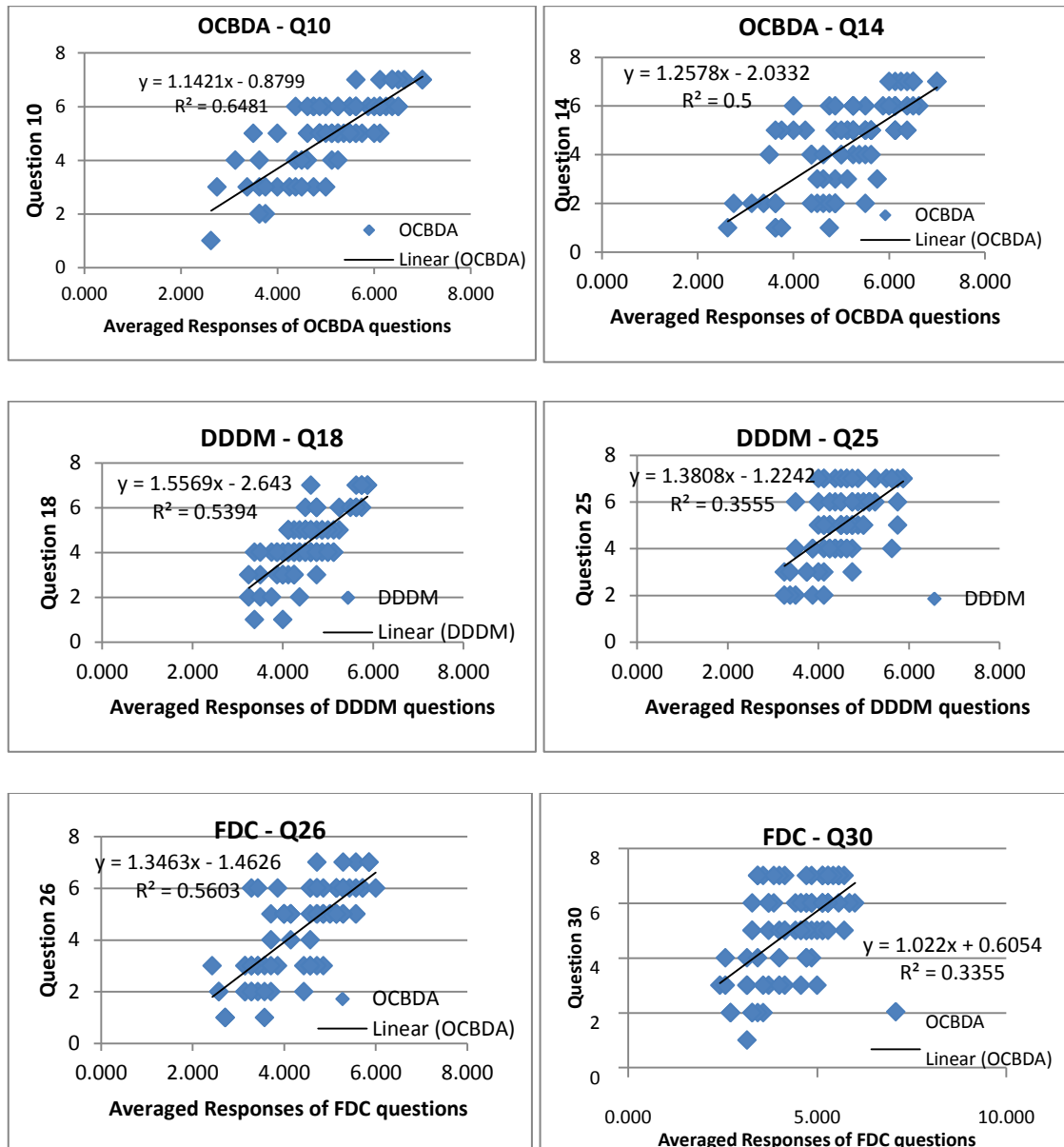
OCBDA	q9	q10	q11	q12	q13	q14	q15	q16	OCBDA									
Pearson Correlation	.472**	.805**	.506**	.755**	.826**	.707**	.755**	.614**	1									
Sig. (2-tailed)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000										
N	80	80	80	80	80	80	80	80	80									
**. Correlation is significant at the 0.01 level (2-tailed).																		
DDDM	q17	q18	q20	q22	q23	q24	DDDM											
Pearson Correlation	.742**	.832**	.412**	.662**	.662**	.778**	1											
Sig. (2-tailed)	0.000	0.000	0.000	0.000	0.000	0.000												
N	80	80	80	80	80	80	80											
**. Correlation is significant at the 0.01 level (2-tailed).																		
FDC	q26	q27	q28	q29	q30	q31	FDC											
Pearson Correlation	.789**	.790**	.884**	.720**	.555**	.718**	1											
Sig. (2-tailed)	0.000	0.000	0.000	0.000	0.000	0.000												
N	80	80	80	80	80	80	80											
**. Correlation is significant at the 0.01 level (2-tailed).																		
BDAMC	q32	q33	q34	q35	q36	q37	q38	q39	q40	q41	q42	q43	q44	q45	q46	q47	q48	BDAMC
Pearson Correlation	.462**	.720**	.580**	.569**	.808**	.670**	.784**	.703**	.693**	.629**	.585**	.709**	.605**	.702**	.497**	.659**	.701**	1
Sig. (2-tailed)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
N	80	80	80	80	80	80	80	80	80	80	80	80	80	80	80	80	80	80
**. Correlation is significant at the 0.01 level (2-tailed).																		
BDAIF	q49	q50	q51	q52	q53	q54	q55	q56	BDAIF									
Pearson Correlation	.676**	.724**	.665**	.785**	.745**	.732**	-.222*	.681**	1									
Sig. (2-tailed)	0.000	0.000	0.000	0.000	0.000	0.000	0.048	0.000										
N	80	80	80	80	80	80	80	80	80									

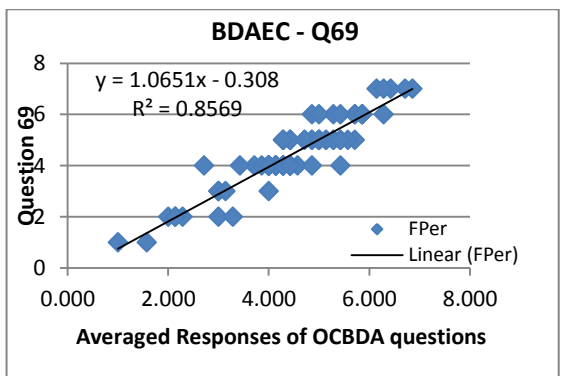
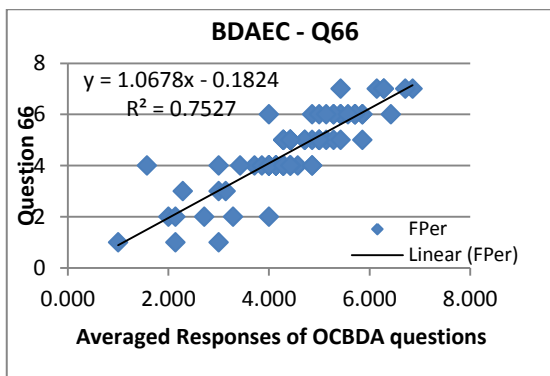
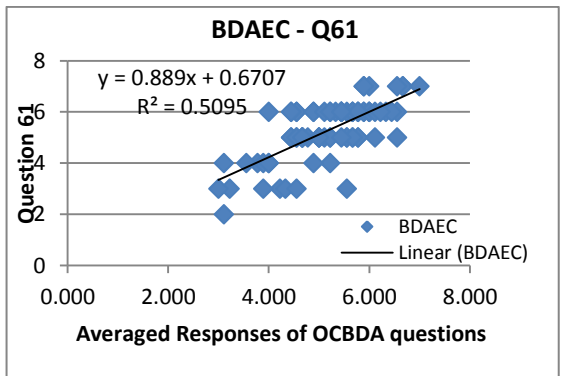
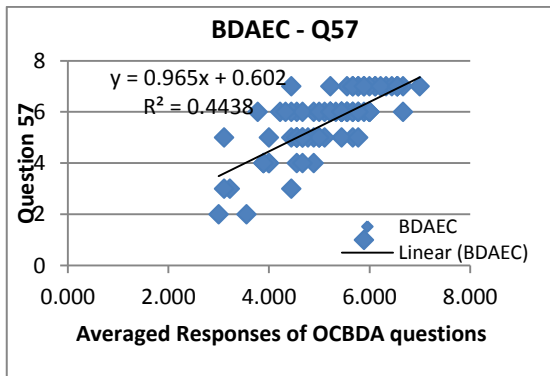
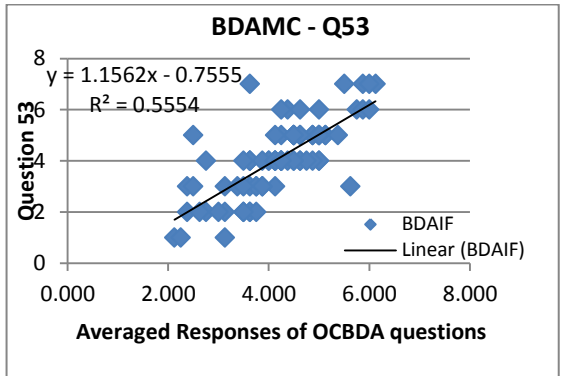
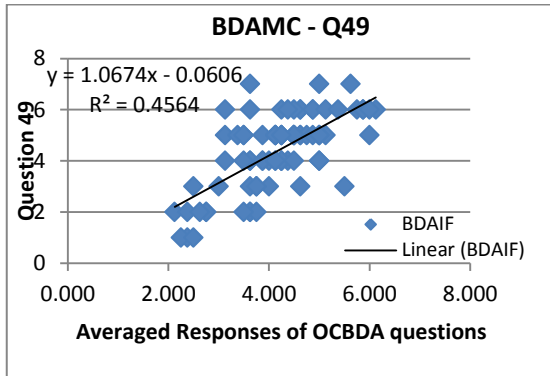
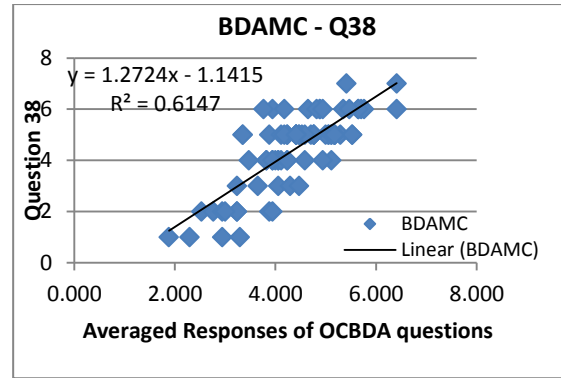
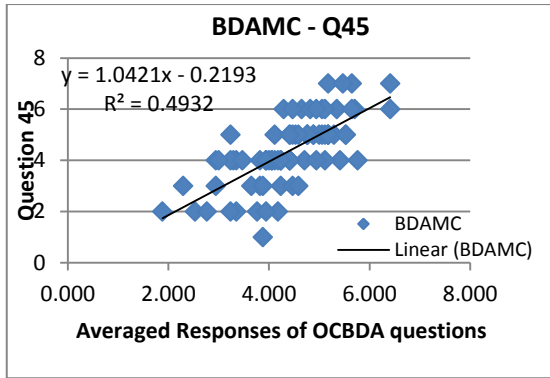
**. Correlation is significant at the 0.01 level (2-tailed).										
*. Correlation is significant at the 0.05 level (2-tailed).										
BDAPE	q57	q58	q59	q60	q61	q62	q63	q64	q65	BDAPE
Pearson Correlation	.666**	.825**	.751**	.804**	.714**	.756**	.685**	.642**	.690**	1
Sig. (2-tailed)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
N	80	80	80	80	80	80	80	80	80	80
**. Correlation is significant at the 0.01 level (2-tailed).										
FPer	q66	q67	q68	q69	q70	q71	q72	Fper		
Pearson Correlation	.868**	.904**	.917**	.926**	.906**	.640**	.755**	1		
Sig. (2-tailed)	0.000	0.000	0.000	0.000	0.000	0.000	0.000			
N	80	80	80	80	80	80	80	80		
**. Correlation is significant at the 0.01 level (2-tailed).										

9.4.4. Assessment linearity between variables by scatter plot for Factor Analysis

In each of the following, sample questions from each construct are presented to illustrate the presence of linearity between variables assessed; such the data can be applied in EFA.

Figure 4: Representative scatter plots illustrating linearity of variables across constructs (FDC, OCBDA and DDDM)





In each case, to allow for a bi-variate comparison, the variables of theoretical constructs as presented from Cronbach's alpha assessment were averaged across each respondent to find the mean of the theoretical construct. These were then set as the x-values of a scatter plot, and the individual question responses per respondent compared to this at the y-variable. In each case, visual linearity as well as a R^2 value in excess of 0.5 was assumed to suggest linearity between variables. As is evident, not all data sets met these criteria, as described by DDDM question 25.

9.4.5. Exploratory factor analysis (EFA)

The raw KMO and Bartlett's test measures for Section 2 (OCBDA) are presented in Table 33 as a representative of the data collected for each construct. Therein, KMO was > 0.7 and Bartlett's was statistically significant, thus the EFA was considered valid, and further assessments conducted.

Table 33: KMO and Bartlett's test of Sections 2 (OCBDA)

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		0.845
Bartlett's Test of Sphericity	232.013	1421.177
	28	406
	0.000	0.000

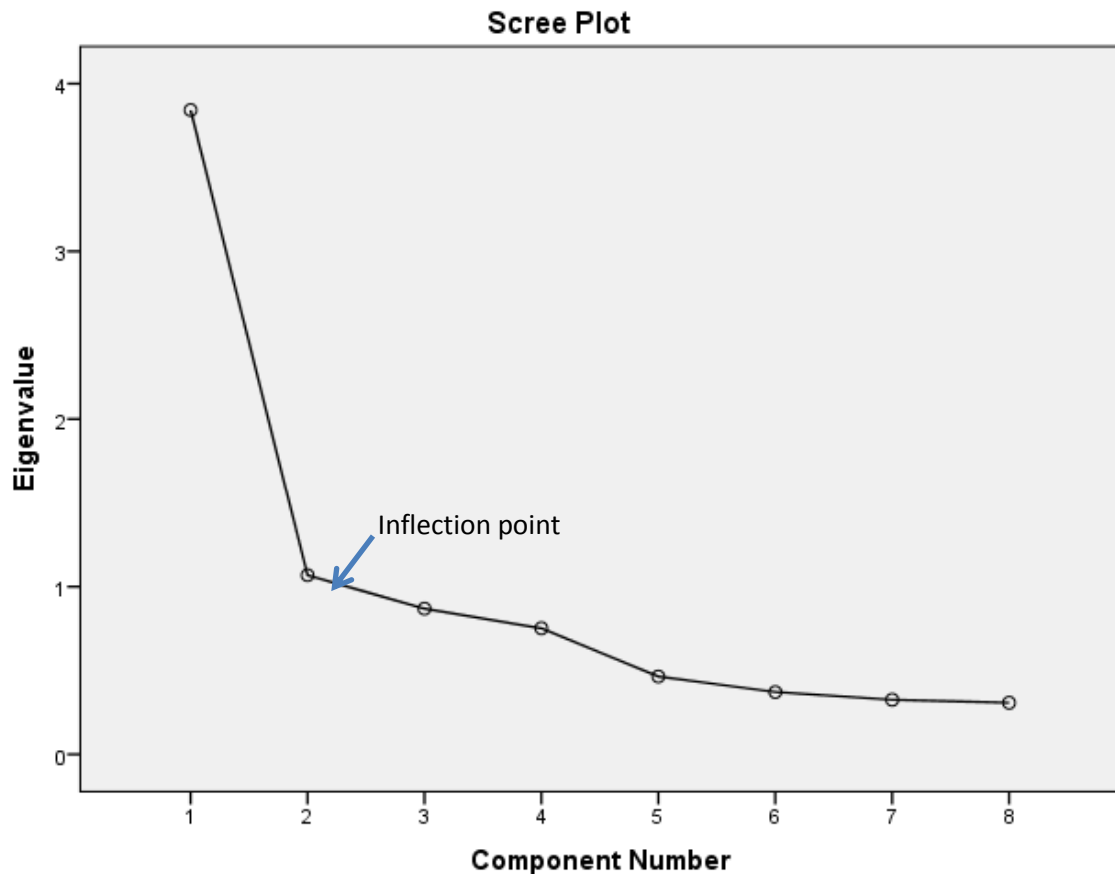
Number of constructs as well as total variance explained by the constructs was illustrated through setting Eigenvalue to 1 (Table 34), as well as through Scree plot (Figure 5).

Table 34: Extraction of number of constructs and total variance explained by Eigenvalues for Section 2 (OCBDA)

Component	Total Variance Explained								
	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3.842	48.030	48.030	3.842	48.030	48.030	2.743	34.293	34.293
2	1.068	13.353	61.383	1.068	13.353	61.383	2.167	27.090	61.383
3	0.869	10.860	72.242						
4	0.752	9.397	81.640						
5	0.464	5.805	87.445						
6	0.372	4.645	92.090						
7	0.325	4.066	96.156						
8	0.308	3.844	100.000						

In both Table 34 and Figure 5, a total of 2 constructs were visibly extracted, with a total variance explained of 61.383%, as illustrated by bold typeface in Table 34. Assessment of a scree plot (Figure 5) revealed an inflection point at 2, which illustrates the number of constructs determined (Costello & Osborne, 2005).

Figure 5: Scree plot of sections 2 (OCBDA)



The constituents or factor loading of the 2 respective constructs are described through Table 35, wherein the rotated matrix illustrates loading by highest value, per row (question), which inform the construct number (column). The results of which component loaded to which construct are highlighted in boldface for ease of reference. In all cases, all question were loaded to a construct, so none were deleted

Table 35: Rotated Component matrix illustrating factor loading per item from EFA for Section 2 (OCBDA)

Rotated Component Matrix ^a		
Question Description	Component	
	1	2
q9_innovation_nb_org_future	0.007	0.695
q10_org_help_learning_app_knowledge	0.766	0.365
q11_adapt_new_practice_continuously	0.148	0.661
q12_org_encourage_staff_new_capabilities	0.581	0.527
q13_org_visionary_flexible	0.559	0.637
q14_employee_orientation_program	0.869	0.017
q15_org_trains_use_of_data	0.814	0.164
q16_exec_support_bda	0.259	0.636
q9_innovation_nb_org_future	0.007	0.695
Extraction Method: Principal Component Analysis.		
Rotation Method: Varimax with Kaiser Normalization.		
a. Rotation converged in 3 iterations.		

9.4.6. Confirmatory factor analysis

Figure 6: CFA model of OCBDA

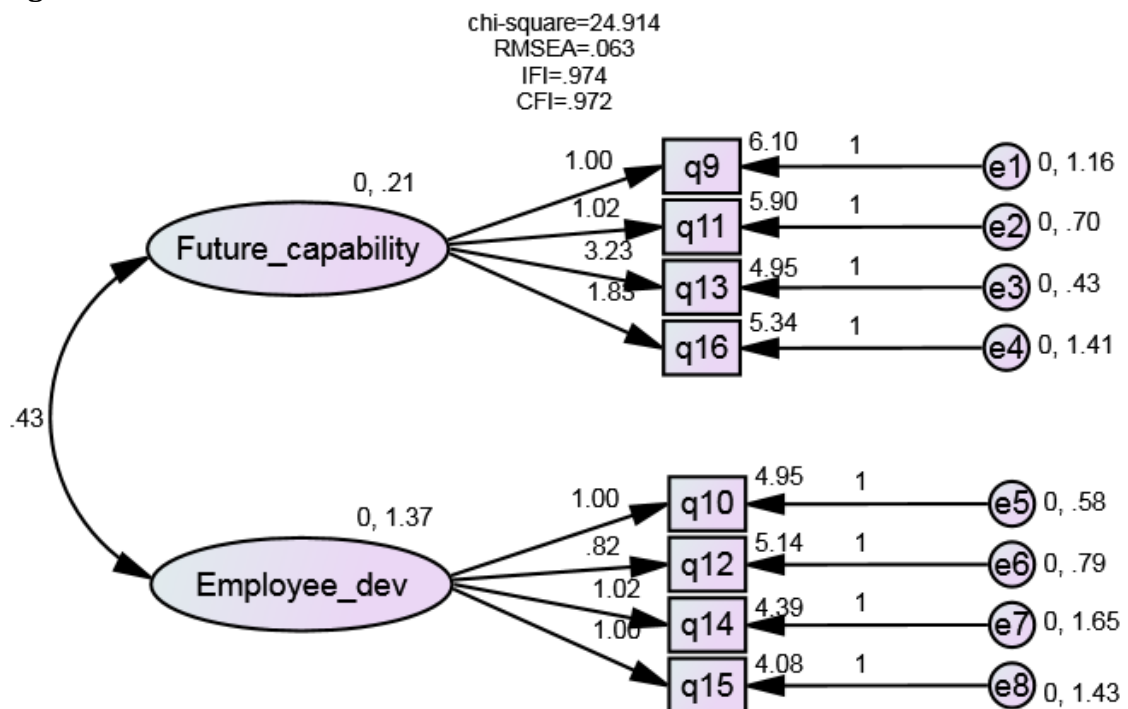


Figure 7: CFA of DDDM

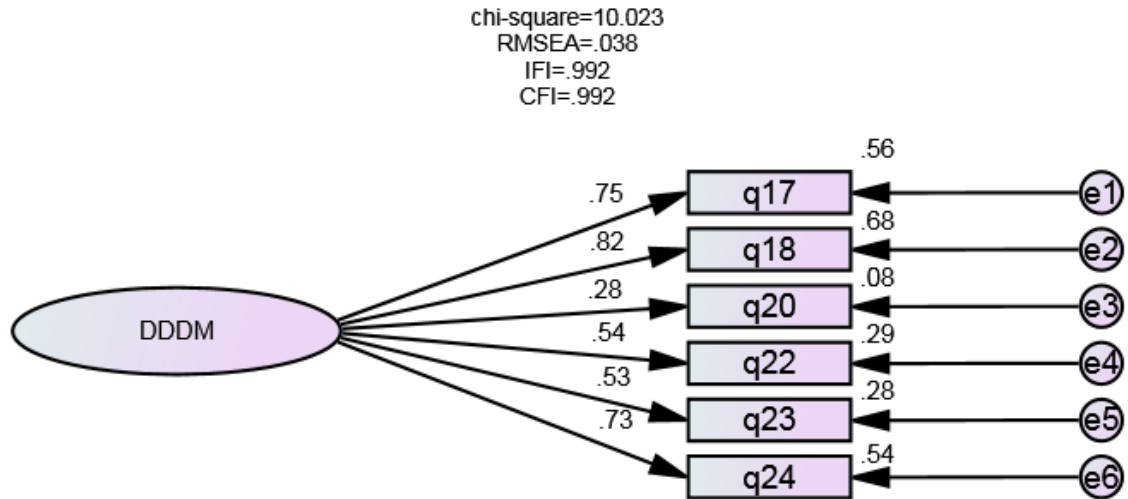


Figure 8: CFA of FDC and FAC

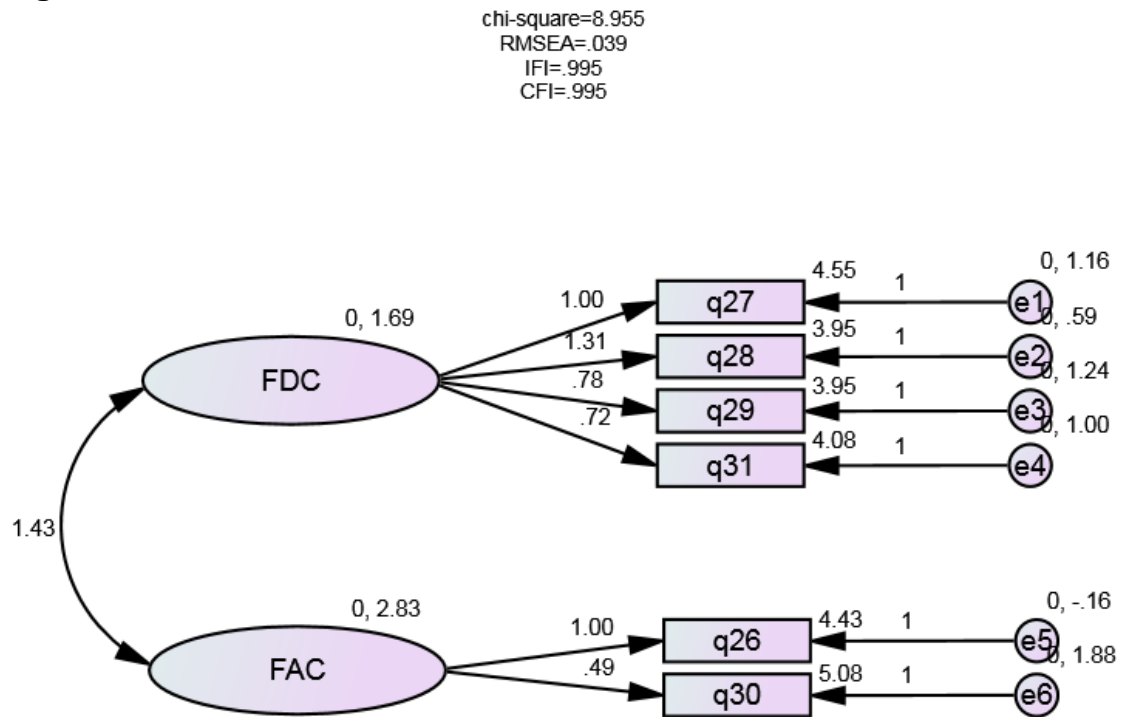


Figure 9: CFA of BDAMC

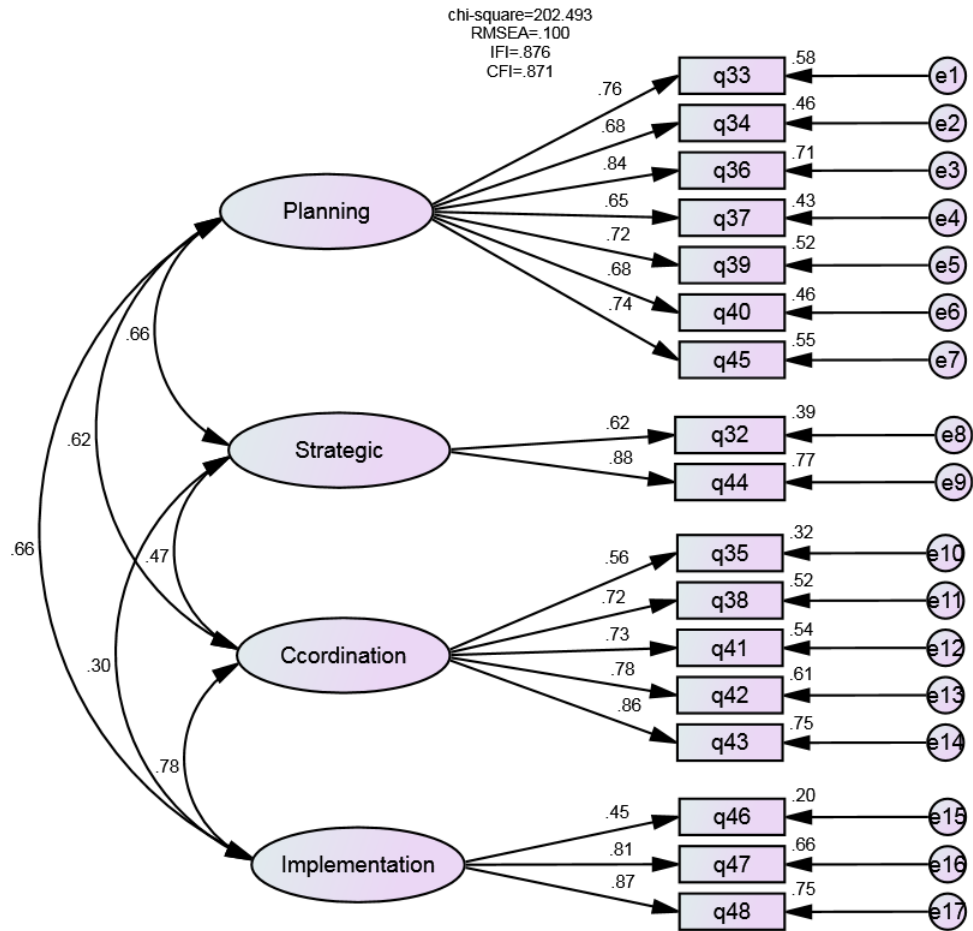


Figure 10: CFA of BDAIC

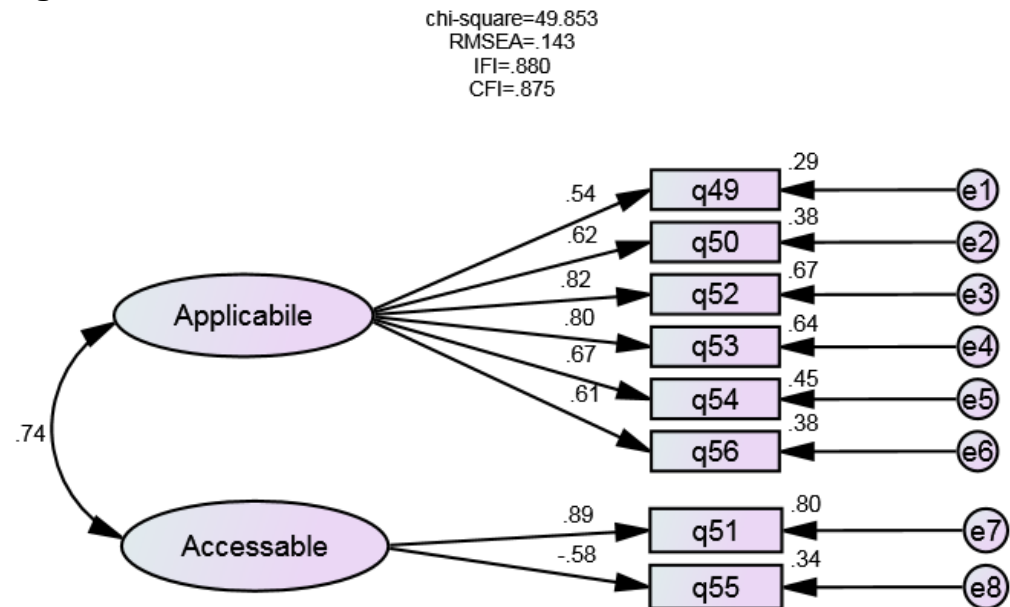


Figure 11: CFA of BDAPE

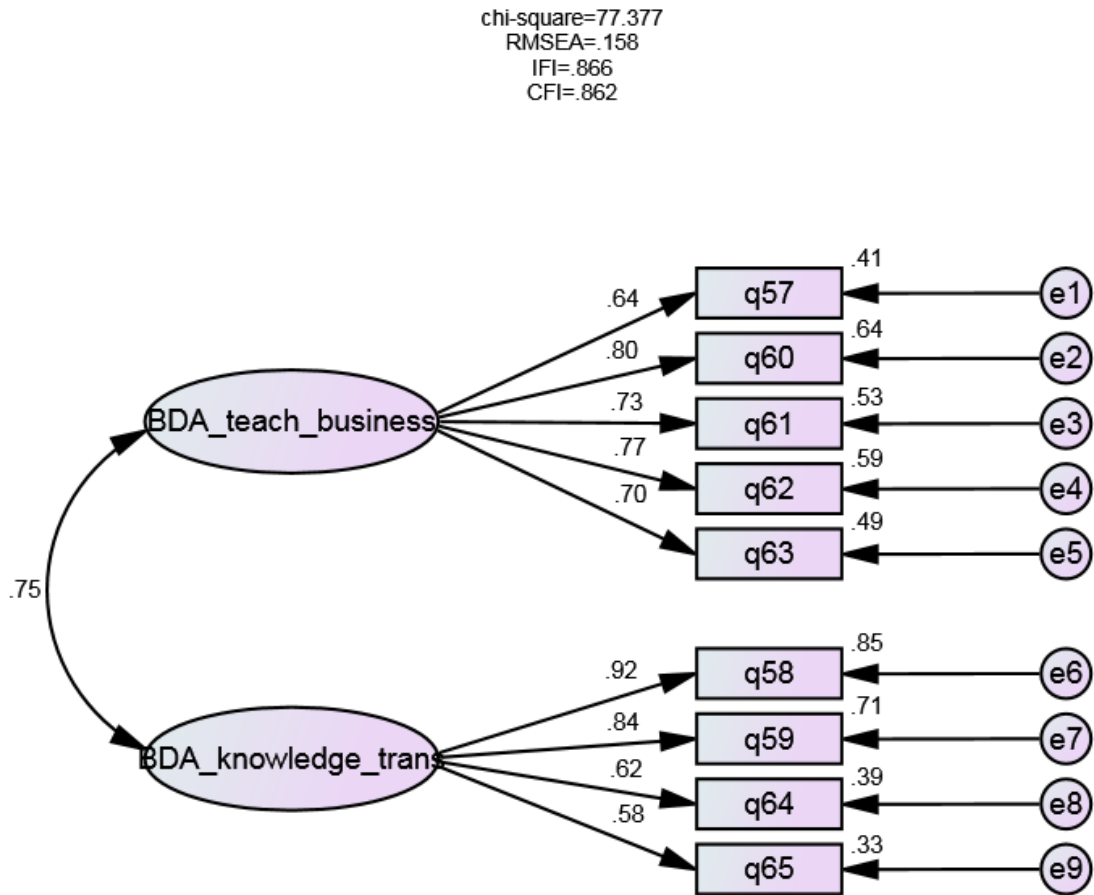


Figure 12: CFA of FPer

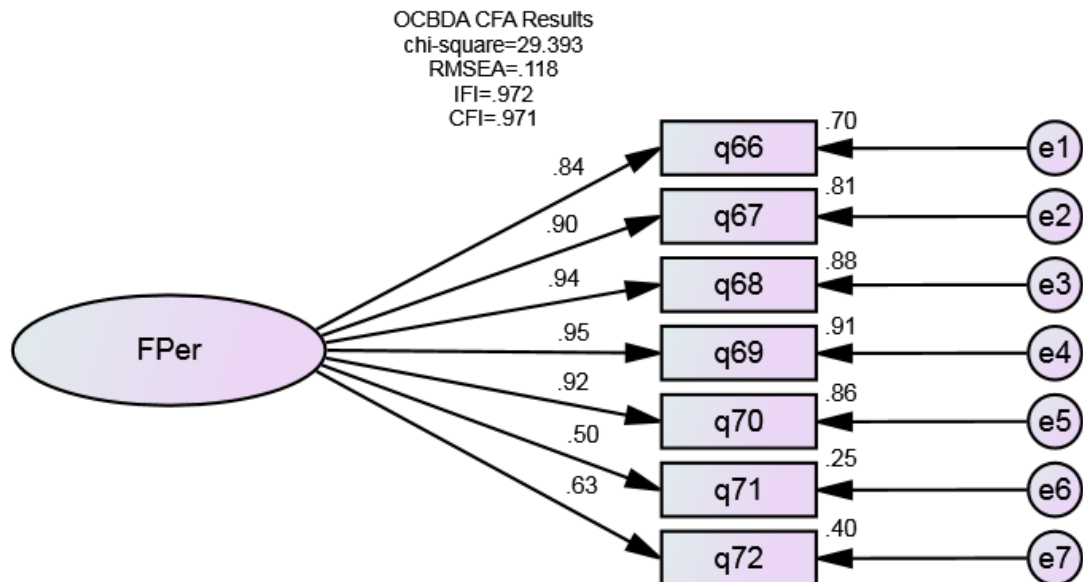


Figure 13: CFA of the third order construct BDAC

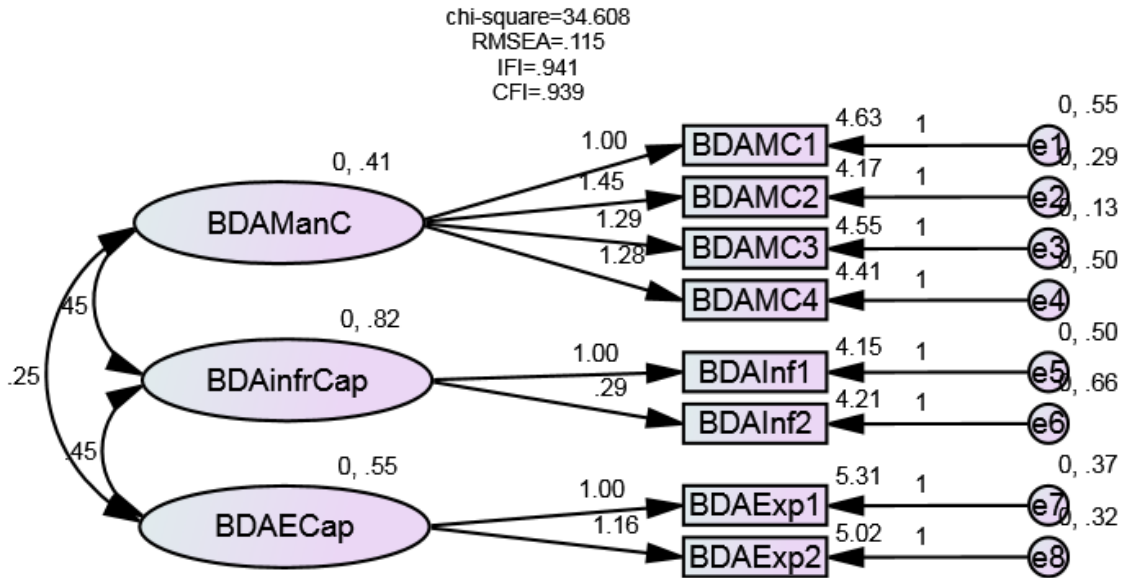
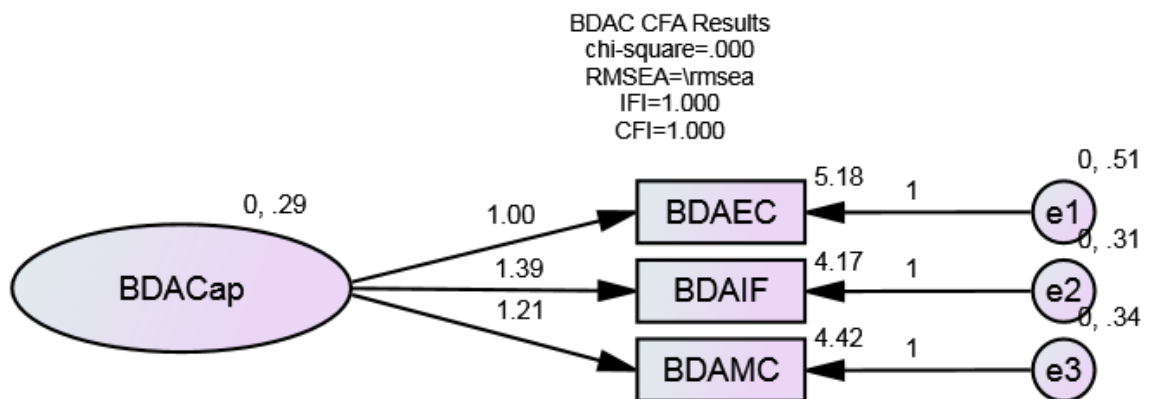


Figure 14: BDAC loading of level 2 constructs



9.4.7. Correlation of Factors

Table 36: Spearman's correlation matrix between all construct factors assessed

Spearman's Rho Correlations		Support Future Capability	Employee Dev.	DDDM	FAC	FDC	Strategic	Coordination	Planning	Implementation	Applicable	Accessible	BDA Tech applicable	BDA Knowledge Transfer	FPer	BDAC
Support Future Capability	Correlation Coefficient	1.000	.608**	.472**	.498**	.606**	.560**	.452**	.528**	.534**	.451**	0.128	.401**	0.214	.391**	.553**
	Sig. (2-tailed)		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.256	0.000	0.057	0.000	0.000
	N	80	80	80	80	80	80	80	80	80	80	80	80	80	80	80
Employee Development	Correlation Coefficient	.608**	1.000	.663**	.472**	.646**	.512**	.592**	.498**	.524**	.410**	0.193	.400**	0.214	.397**	.588**
	Sig. (2-tailed)	0.000		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.086	0.000	0.056	0.000	0.000
	N	80	80	80	80	80	80	80	80	80	80	80	80	80	80	80
DDDM	Correlation Coefficient	.472**	.663**	1.000	.361**	.643**	.533**	.660**	.622**	.456**	.460**	.223*	.388**	.269*	.524**	.657**
	Sig. (2-tailed)	0.000	0.000		0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.047	0.000	0.016	0.000	0.000
	N	80	80	80	80	80	80	80	80	80	80	80	80	80	80	80
FAC	Correlation Coefficient	.498**	.472**	.361**	1.000	.518**	.353**	.325**	.273*	0.203	.420**	0.039	0.157	0.069	.318**	.306**
	Sig. (2-tailed)	0.000	0.000	0.001		0.000	0.001	0.003	0.014	0.071	0.000	0.728	0.164	0.545	0.004	0.006
	N	80	80	80	80	80	80	80	80	80	80	80	80	80	80	80
FDC	Correlation Coefficient	.606**	.646**	.643**	.518**	1.000	.591**	.573**	.657**	.531**	.600**	0.120	.310**	0.219	.636**	.652**
	Sig. (2-tailed)	0.000	0.000	0.000	0.000		0.000	0.000	0.000	0.000	0.000	0.288	0.005	0.051	0.000	0.000
	N	80	80	80	80	80	80	80	80	80	80	80	80	80	80	80
Strategic	Correlation	.560**	.512**	.533**	.353**	.591**	1.000	.534**	.677**	.536**	.478**	0.144	.236*	0.194	.436**	.613**

	Coefficient															
	Sig. (2-tailed)	0.000	0.000	0.000	0.001	0.000		0.000	0.000	0.000	0.000	0.204	0.035	0.084	0.000	0.000
	N	80	80	80	80	80	80	80	80	80	80	80	80	80	80	80
Coordination	Correlation Coefficient	.452**	.592**	.660**	.325**	.573**	.534**	1.000	.771**	.669**	.497**	0.220	.308**	.307**	.621**	.773**
	Sig. (2-tailed)	0.000	0.000	0.000	0.003	0.000	0.000		0.000	0.000	0.000	0.050	0.005	0.006	0.000	0.000
	N	80	80	80	80	80	80	80	80	80	80	80	80	80	80	80
Planning	Correlation Coefficient	.528**	.498**	.622**	.273*	.657**	.677**	.771**	1.000	.696**	.533**	0.128	.409**	.389**	.648**	.831**
	Sig. (2-tailed)	0.000	0.000	0.000	0.014	0.000	0.000	0.000		0.000	0.000	0.257	0.000	0.000	0.000	0.000
	N	80	80	80	80	80	80	80	80	80	80	80	80	80	80	80
Implementation	Correlation Coefficient	.534**	.524**	.456**	0.203	.531**	.536**	.669**	.696**	1.000	.659**	.272*	.446**	.445**	.553**	.831**
	Sig. (2-tailed)	0.000	0.000	0.000	0.071	0.000	0.000	0.000	0.000		0.000	0.014	0.000	0.000	0.000	0.000
	N	80	80	80	80	80	80	80	80	80	80	80	80	80	80	80
Applicability	Correlation Coefficient	.451**	.410**	.460**	.420**	.600**	.478**	.497**	.533**	.659**	1.000	.250*	.387**	.456**	.516**	.783**
	Sig. (2-tailed)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000		0.025	0.000	0.000	0.000	0.000
	N	80	80	80	80	80	80	80	80	80	80	80	80	80	80	80
Accessibility	Correlation Coefficient	0.128	0.193	.223*	0.039	0.120	0.144	0.220	0.128	.272*	.250*	1.000	.265*	.237*	0.051	.340**
	Sig. (2-tailed)	0.256	0.086	0.047	0.728	0.288	0.204	0.050	0.257	0.014	0.025		0.018	0.034	0.654	0.002
	N	80	80	80	80	80	80	80	80	80	80	80	80	80	80	80
BDA Tech applicable to business	Correlation Coefficient	.401**	.400**	.388**	0.157	.310**	.236*	.308**	.409**	.446**	.387**	.265*	1.000	.744**	.341**	.638**
	Sig. (2-tailed)	0.000	0.000	0.000	0.164	0.005	0.035	0.005	0.000	0.000	0.000	0.018		0.000	0.002	0.000

	N	80	80	80	80	80	80	80	80	80	80	80	80	80	80	80
BDA Knowledge Transfer	Correlation Coefficient	0.214	0.214	.269 *	0.069	0.219	0.194	.307 **	.389 **	.445 **	.456 **	.237 *	.744 **	1.000	.305 **	.658 **
	Sig. (2-tailed)	0.057	0.056	0.016	0.545	0.051	0.084	0.006	0.000	0.000	0.000	0.034	0.000		0.006	0.000
	N	80	80	80	80	80	80	80	80	80	80	80	80	80	80	80
Firm Performance	Correlation Coefficient	.391 **	.397 **	.524 **	.318 **	.636 **	.436 **	.621 **	.648 **	.553 **	.516 **	0.051	.341 **	.305 **	1.000	.632 **
	Sig. (2-tailed)	0.000	0.000	0.000	0.004	0.000	0.000	0.000	0.000	0.000	0.000	0.654	0.002	0.006		0.000
	N	80	80	80	80	80	80	80	80	80	80	80	80	80	80	80
BDAC	Correlation Coefficient	.553 **	.588 **	.657 **	.306 **	.652 **	.613 **	.773 **	.831 **	.831 **	.783 **	.340 **	.638 **	.658 **	.632 **	1.000
	Sig. (2-tailed)	0.000	0.000	0.000	0.006	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.000	0.000	0.000	
	N	80	80	80	80	80	80	80	80	80	80	80	80	80	80	80

Table 37: Regression of 2nd order BDAC constructs to BDAC (3rd order)

		BDAC (3 rd Order)	BDAMC (2 nd Order)	BDAIF (2 nd Order)	BDAEC (2 nd Order)
Pearson Correlation	BDAC	1.000	.907	.799	.724
	BDAMC	.907	1.000	.596	.448
	BDAIF	.799	.596	1.000	.481
	BDAEC	.724	.448	.481	1.000
Sig. (1-tailed)	BDAC	.	.000	.000	.000
	BDAMC	.000	.	.000	.000
	BDAIF	.000	.000	.	.000
	BDAEC	.000	.000	.000	.
N	BDAC	80	80	80	80
	BDAMC	80	80	80	80
	BDAIF	80	80	80	80
	BDAEC	80	80	80	80

9.4.8. Scatter plots of each factor (Y) and FPer (X)

Figure 15: Scatter plot of each factor with FPer to illustrate relative distribution

