T.R.
SAKARYA UNIVERSITY
GRADUATE SCHOOL OF BUSINESS

CRITICAL SUCCESS FACTORS OF BIG DATA PROJECTS: A MODEL PROPOSAL AND EMPIRICAL TEST

DOCTORAL THESIS

Naciye Güliz UĞUR

Department: Management Information Systems

Supervisor: Prof. Dr. Aykut Hamit TURAN

DECEMBER – 2018

T.R.
SAKARYA UNIVERSITY
GRADUATE SCHOOL OF BUSINESS

CRITICAL SUCCESS FACTORS OF BIG DATA PROJECTS: A MODEL PROPOSAL AND EMPIRICAL TEST

DOCTORAL THESIS

Naciye Güliz UĞUR

Department: Management Information Systems

"This thesis has been accepted unanimously / with majority of vote by the following jury on 28/12/2018"

JURY MEMBER	DECISION	SIGNATURE
Prof. Dr. Aykut Hamit Turan	Accepted	
PROF. DR. B. AKKUT ARIKAN	ACCEPTED	
Prof. Dr. ERMAN COŞKUN	ACCEPTED	
PROF. DR. BİRGÜL KUTLU BAYRAKTAR	ACCEPTED	
Doç. Dr. Mustafa C Ungan	Accepted	

SAKARYA ÜNİVERSİTESİ	T.C. SAKARYA ÜNİVERSİTESİ İŞLETME ENSTİTÜSÜ TEZ SAVUNULABİLİRLİK VE ORJİNALLİK BEYAN FORMU	Sayfa : 1/1

Öğrencinin

Adı Soyadı	:	NACİYE GÜLİZ UĞUR
Öğrenci Numarası	:	D156054001
Enstitü Anabilim Dalı	:	YÖNETİM BİLİŞİM SİSTEMLERİ
Enstitü Bilim Dalı	:	
Programı	:	☐ YÜKSEK LİSANS ☑ DOKTORA
Tezin Başlığı	:	ORGANIZATIONAL EFFECTS OF BIG DATA: A RESEARCH ON CRITICAL SUCCESS FACTORS
Benzerlik Oranı	:	% 11

ENSTİTÜSÜ MÜDÜRLÜĞÜNE

☑ Sakarya Üniversitesi İşletme Enstitüsü Lisansüstü Tez Çalışması Benzerlik Raporu Uygulama Esaslarını inceledim. Enstitünüz tarafından Uygulalma Esasları çerçevesinde alınan Benzerlik Raporuna göre yukarıda bilgileri verilen tez çalışmasının benzerlik oranının herhangi bir intihal içermediğini; aksinin tespit edileceği muhtemel durumda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi beyan ederim.

10/12/2018
İmza

☐ Sakarya Üniversitesi İşletme Enstitüsü Lisansüstü Tez Çalışması Benzerlik Raporu Uygulama Esaslarını inceledim. Enstitünüz tarafından Uygulalma Esasları çerçevesinde alınan Benzerlik Raporuna göre yukarıda bilgileri verilen öğrenciye ait tez çalışması ile ilgili gerekli düzenleme tarafımca yapılmış olup, yeniden değerlendirilmek üzere gsbtez@sakarya.edu.tr adresine yüklenmiştir.

Bilgilerinize arz ederim.

Uygundur

Danışman
Unvanı / Adı-Soyadı: PROF. DR. AYKUT HAMİT TURAN

Tarih: 10/12/2018

İmza:

☐ KABUL EDİLMİŞTİR ☐ REDDEDİLMİŞTİR **EYK Tarih ve No:**	**Enstitü Birim Sorumlusu Onayı**

00.ENS.FR.72

PREFACE

"When eating a fruit, think of the person who planted the tree".

Vietnamese Proverb

Embarking on a Ph.D. project is a remarkable undertaking. This journey was made possible by the help and guidance of some precious people, who have provided advice, encouragement, and support to me in my progress toward the completion of this journey.

First and foremost, I would like to express my deepest appreciation to my advisor, Prof. Dr. Aykut Hamit Turan, for his academic guidance and patient, friendly, and unfailing support over the past six years. I am sincerely indebted to Prof. Dr. Erman Coşkun, who started my academic career by enabling me to be a research assistant. I am grateful for his endless support and contribution to my work. I am thankful to Assoc. Prof. Dr. Mustafa Cahid Ünğan for his critical suggestions and advice in various thesis progress meetings that set the direction and focus for my research. Every meeting with them proved significant in advancing my dissertation. I am deeply grateful to Assist. Prof. Dr. Adem Akbıyık for his countless contribution to my dissertation, I am inspired by his methodological knowledge and perfectionism. I would also like to thank my committee members, Prof. Dr. Birgül Kutlu Bayraktar and Prof. Dr. Aykut Arıkan for their valuable contribution, I am honored to be approved by these precious professors and to have their signatures in my dissertation.

I am thankful to my oldest instructor Prof. Dr. Alptekin Erkollar, who have been present in all stages of my higher education and I owe a great debt to Prof. Dr. Yılmaz Gökşen for his valuable contribution to my doctoral qualification.

I am sincerely indebted to my dearest aunt-in-law Prof. Dr. Ayşen Özel who instilled in me in my childhood the dream of becoming an academician and I am deeply grateful to Assist. Dr. Esin Cevrioğlu, who gave me the greatest support for realizing my dream and becoming a research assistant.

Finally, my gratitude is endless to my beloved husband Emre Uğur and my honeybunch daughter Duru Bilge Uğur for their unlimited support and patience throughout my academic life. Without your perseverance, this degree would not be possible.

This dissertation is dedicated to the memory of my nearest and dearest mother and granny, even though you are not with me physically, I always feel your endless love and prayer. Rest in peace.

I thank all of you for planting this tree…

Naciye Güliz UĞUR

28. 12. 2018

TABLE OF CONTENTS

ABBREVIATIONS

AMOS : Analysis of Moments Structure
AVE : Average Variance Extracted
BDA : Big Data Analytics
BI : Business Intelligence
CATI : Computer-Assisted Telephone Interview
CB : Covariance Based
CEO : Chief Executive Officer
CFA : Confirmatory Factor Analysis
CGT : Classic Grounded Theory
CMMI : Capability Maturity Model Integration
CMV : Common Method Variance
CPU : Central Processing Unit
CSF : Critical Success Factor
DTPB : Decomposed Theory of Planned Behavior
DB : Database
EFA : Exploratory Factor Analysis
ERP : Enterprise Resource Planning
ETL : Extract – Transform - Load
G : Governance (construct)
HCM : Hierarchical Component Models
HOC : Higher-Order Construct
HTML : HyperText Markup Language
HTMT : Heterotrait-Monotrait Ratio
IDC : International Data Corporation
IDT : Innovation Diffusion Theory
IQR : Interquartile Range
IS : Information System
IT : Information Technology
IoT : Internet of Things
K-S : Kolmogorov-Smirnov
KMO : Kaiser-Meyer-Olkin
LOC : Lower-Order Construct
MC : Multicollinearity
MIS : Management Information Systems
ML : Machine Learning
MM : Motivation Model
MPCU : Model of PC Utilization
MRA : Multiple Regression Analysis
PCA : Principal Component Analysis
PD : Project Definition (construct)
PLS : Partial Least Square
PM : Project Management (construct)
PMBOK : Project Management Book of Knowledge
PMI : Project Management Institute
PPRL : Privacy-Preserving Record Linkage

RFID : Radio Frequency Identification
S : Success (construct)
SAP : Systems Analysis and Program Development
SCT : Social Cognitive Theory
SDLC : System Development Life Cycle
SEM : Structural Equation Modeling
SOA : Service-Oriented Architecture
SPSS : Statistical Package for the Social Sciences
SKU : Stock-Keeping Unit
SRMR : Standardized Root Mean Square Residual
TAM : Technology Acceptance Model
TC : Technology (construct)
TM : Team (construct)
TPB : Theory of Planned Behavior
TRA : Theory of Reasoned Action
UTAUT : Unified Theory of Acceptance and Use of Technology
VIF : Variance Inflation Factor
WoS : Web of Science
XML : Extensible Markup Language

LIST OF TABLES

LIST OF FIGURES

Title of the Thesis: Critical Success Factors of Big Data Projects: A Model Proposal and Empirical Test	
Author: Naciye Güliz UĞUR	**Supervisor:** Prof. Dr. Aykut Hamit TURAN
Date: 28 December 2018	**Nu. of pages:** xii (pre text) + 265 (main body) + 6 (App.)
Department: Management Information Systems	

The explosion of data being captured and stored in information systems has created a new area of challenges and opportunities for information technology (IT) professionals. While substantial efforts have been made towards algorithms and technologies that are used to perform these analytics, comparatively there has been limited empirical research on Critical Success Factors (CSFs) that relate to Big Data projects.

The lack of critical success factor sources can doom an IS project to a certain failure. This research promises to help organizations to identify factors that impact success – as perceived by practitioners and professionals – on Big Data projects.

The main purpose of this research is to build on the current diverse literature around Big Data by contributing discussion and data that allow common agreement on factors that influence successful Big Data projects. The research also validates the CSF scale and theoretical CSF model statistically. While individual and technical factors have been explored as they relate to Big Data success, there is a gap in the literature in determining the critical factors in the light of the views of Big Data experts. Even though critical success factors have been discussed previously as being related to IS success, it has not been associated with Big Data project success. The most complete information regarding the CSFs for Big Data projects can be received from Big Data professionals within those departments that have been involved in Big Data projects. Accordingly, this study is conducted with 17 Big Data experts in earlier Delphi Study and 827 Big Data professionals in large scale survey administration. At the end of the study, five CSFs emerged in addition to a statistically reliable and valid CSF measurement scale and a relational research model that is tested and validated.

This research is exploratory in nature. The best approach for such a study was mixed methods utilizing Constructivist Grounded Theory. Grounded theory allows the researcher to begin with the question, collect data, examine ideas and concepts, extract and categorize that data to use it to form the basis of a new theory. This new theory can then be applied and tested statistically. To successfully accomplish this, the approach for the study was fragmented into a three-part mixed methods study. A qualitative section utilizing semi-structured interviews and Delphi study with experts in the field followed by a quantitative section to test relationships between core concepts derived from the qualitative section.

Keywords: critical success factors, big data, scale development, Delphi study, empirical study

Tezin Başlığı: Büyük Veri Projelerinin Kritik Başarı Faktörleri: Bir Model Önerisi ve Ampirik Test

Tezin Yazarı: Naciye Güliz UĞUR **Danışman:** Prof. Dr. Aykut Hamit

Kabul Tarihi: 28 Aralık 2018 **Sayfa Sayısı:** xii (ön kısım) + 265 (tez) + 6 (ek)

Anabilimdalı: Yönetim Bilişim Sistemleri

Bilişim sistemleri vasıtasıyla elde edilen ve depolanan verilerin hızla artması, bilişim teknolojisi (BT) uzmanları için yeni zorlukları ve fırsatları beraberinde getirmiştir. Veri analitiği bağlamında kullanılan algoritma ve teknolojilere yönelik önemli çabalar gösterilmiş olmasına karşın, Büyük Veri projelerine yönelik Kritik Başarı Faktörleri (KBF'ler) üzerine yapılan araştırmalar sınırlı sayıdadır. Kritik başarı faktörleri daha önce Bilişim Sistemleri alanında tartışılmış olmakta beraber, bulgular Büyük Veri projeleri ile ilişkilendirilmemiştir.

Kritik başarı faktörü kaynaklarının eksikliği, bir Bilişim Sistemi projesini başarısızlığa mahkum edebilir. Bu araştırma, işletmelerin Büyük Veri projelerinin başarısını etkileyen kritik faktörleri tespit etmelerine yardımcı olmayı vaat etmektedir.

Bu araştırmanın temel amacı, Büyük Veri projelerini etkileyen başarı faktörleri üzerinde anlaşmaya varılmasına olanak tanıyan tartışmaya ve verilere katkıda bulunarak, Büyük Veri odaklı literature katkı sağlamaktır. Bunun yanı sıra, araştırma kapsamında KBF ölçeği ve ilişkisel KBF modeli istatistiksel olarak test edilmiş ve doğrulanmıştır. Büyük Veri başarısı ile ilgili araştırmalarda bireysel ve teknik faktörler incelenmiş olsa da, daha geniş bir alanı kapsayan kritik faktörlerin Büyük Veri uzmanlarının görüşleri ışığında belirlenmesine ilişkin bir boşluk bulunmaktadır. Büyük Veri projeleri için KBF'ler ile ilgili en kapsamlı bilgi, Büyük Veri projelerine katkı sağlayan departmanlarda istihdam edilen Büyük Veri profesyonellerinden alınabilir. B bağlamda, bu çalışma 17 Büyük Veri uzmanının katkıları ve 827 Büyük Veri profesyonelinin katılımı ile gerçekleştirilmiştir. Çalışmanın sonunda beş KBF ortaya çıkartılmış ve istatistiksel olarak güvenilir ve geçerli bir ölçek ile %51,8 açıklama gücü olan ilişkisel bir araştırma modeli literature eklenmiştir.

Bu araştırma doğası gereği keşifseldir. Böyle bir çalışma için en uygun yaklaşım Yapısal Gömülü Teori ve beraberinde karma yöntem olarak belirlenmiştir. Gömülü teori, araştırmacının bir soruyla yola çıkmasına, veri toplamasına, fikirleri ve kavramları incelemesine, bu verileri yeni bir teorinin temelini oluşturmak için kullanmasına, bunları ayıklamasına ve kategorize etmesine olanak sağlar. Sonrasında bu yeni teori uygulanabilir ve istatistiksel olarak test edilebilir. Bunu başarılı bir şekilde gerçekleştirmek için, çalışma yaklaşımı üç bölümlü karma yöntem olarak parçalara ayrılmıştır. Yarı yapılandırılmış mülakatlardan ve alandaki uzmanlarla yapılan Delphi uygulamasından oluşan nitel bölümü, nitel bölümden türetilen temel kavramlar arasındaki ilişkileri test etmek için kurgulanan nicel bölüm takip etmektedir.

Anahtar Kelimeler: kritik başarı faktörleri, büyük veri, ölçek geliştirme, araştırma modeli, ampirik çalışma

INTRODUCTION

"Torture the data, and it will confess to anything."

- Ronald Coase, British economist and author, 1977

About 400 years ago, Galileo observed that "the book of nature is written in the mathematics language". This evaluation is still appropriate today given the enormous amount of data sources and actual volume of data being available (McAfee and Brynjolfsson 2012; Jagadish et al 2014; Manyika et al., 2011; Kiron and Shockley, 2011). Simple online platforms and technological advances have made accessibility of data as a reality. This explosion of sources benefits us with gleaning knowledge, insights and opportunities (Xu et al 2015; Chen et al 2012; Forrester 2012; Wamba et al., 2015). On the one hand, the collection, analysis, and amalgamation of this data is creating challenges and questioning current practices, ethics, procedures, and processes (Mantelero and Vaciago, 2015; McAfee and Brynjolfsson, 2012; Gudivada et al., 2015; Punathambekar and Kavada 2015), on the other hand, it creates new opportunities as novel business streams (Wamba et al., 2015). One such business stream, however, deals with organizations realizing their own value of data housed and shared to create information-based products and services for transactional (profits/money) or strategic value of some kind (Wixom et al., 2014).

Due to advancements in technology like cloud computing, internet of things, social networking devices and more, use of mobile-applications is now generating greater quantities of data than ever before. According to the technology research firm Gartner, there will be 25 billion network-connected devices by 2020 (Vass, 2016). However, due to the huge volume of data generated, the high velocity, with which new data are arriving, and the large variety of heterogeneous data, the current quality of data is far from perfect (IDC, 2013). To put Big Data into perspective, roughly ~2.5 exabytes of data is being created every day and that figure is doubling every 40 months (McAfee and Brynjolfsson, 2012). Similarly, other reports, like Halaweh and Massry (2015) estimate ~5 exabytes of data created every two days and a grand total of 8 Zettabytes by 2015 (the equivalent of 18 million Libraries of Congress), which is consistent with McAfee and Brynjolfsson's findings.

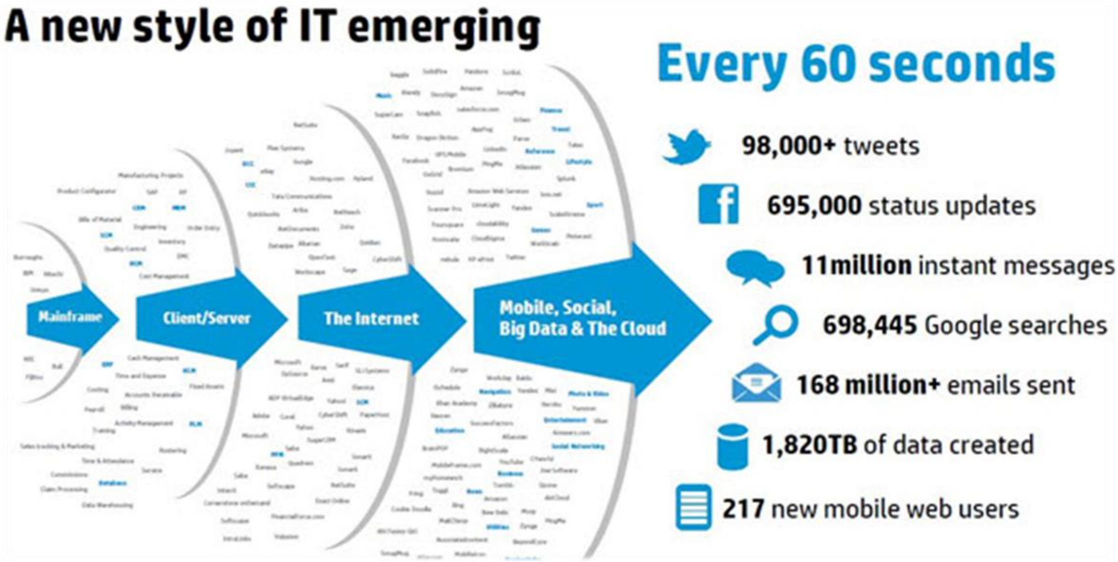

Figure 1: 60 Seconds Statistics

Source: DBTalks, 2016

There are more stats available from various other internet applications such as Amazon.com, Snapchat, Skype, iTunes, Twitter and Pinterest that further highlight the variety of this voluminous data. Keep in mind that this time-box data capture is not restricted only to Internet-ready applications or specific industries. Virtually all industries have their own variations and mechanisms of data collection process, use and value creation of products/services. Industries such as technology, education, healthcare, insurance, finance/banking, commerce and even retail are investigating how they can increase the amount of data that they collect, possess and use.

Every day millions upon millions of bytes of data are being collected, as related to customer transactions, social media postings, government operations, and traffic sensors. The advent of this rise in data presents challenges from technical, managerial, and analytical perspectives. Organizations are being faced with difficult decisions related to the retention of data and how to analyze and stored data to extract value. If organizations hope to obtain value from big data, they must understand the breadth and depth of big data awareness held by their IT employees.

Introduction to the Problem

As the ability to collect, store, and analyze an ever-increasing amount of data generated with a growing frequency, Big Data is a rapidly advancing field. The explosion of data being captured and stored in information systems has created a new area of challenges and opportunities for information technology (IT) professionals. While substantial efforts have been made towards algorithms and technologies that are used to perform the analytics, comparatively fewer efforts have been done toward determining how the organization should work to complete a Big Data project successfully (Saltz and Shamshurin, 2016). Organizations tackling Big Data need more than just knowledge of analytics; they also need the capacity to manage effectively the Big Data effort.

There has been limited empirical research on organizational factors that relate to Big Data (LaValle et al., 2011; Bean and Kiron, 2013). Even though there has been some empirical work on the technical, organizational, and individual factors related to Big Data adoption and success (Uğur and Turan, 2018; Al-Qirim et al., 2017), a gap exists in terms of understanding the critical success factors (CSFs), such as organizational size and top management support, that relate to Big Data project's success. Previous studies have focused primarily on the technical and individual issues relating to Big Data adoption. Sim (2014) acknowledged this gap and suggested that organizations should be aware of the important factors for Big Data success.

Critical success factors have not been investigated as a group of organizational factors that relates to Big Data success. However, researchers have examined critical success factors as an important factor during IS implementations (Davis, 2014; Dong, 2008; Tarhini, Ammar, Tarhini, and Masa'deh, 2015). Several authors have conducted quantitative studies of how critical success factors support relates to specific technologies, including service-oriented architecture (SOA) (Maclennan and Van Belle, 2014), accounting information systems (Anggadini, 2015), healthcare information systems (Hung et al., 2014), and Enterprise Resource Planning (ERP) systems (Dong, Neufeld, and Higgins, 2009; Palanisamy , 2010; Tarhini et al., 2015).

The lack of critical success factor sources can doom an IS project to certain failure. Elbanna (2013) argued that critical success factors have to be consistent and perpetual during a project implementation, otherwise the project would fail. Although IS success was studied in IS implementation process, critical success factors have not been discussed in Big Data projects. Some critical success factors are significant for both IS projects and also for Big Data projects. Top management support is one of these common critical success factors (Barclay, 2015; 2016; Young and Poon, 2013). Young and Poon (2013) suggested that top management support is nearly always necessary for an IS project to be successful because the top management team can influence the success or failure of a project. Conversely, Young and Jordan (2008) argued that project planning, user involvement, and project methodology are not critical success factors for an IS project. But these factors may be critical for a Big Data project. Big Data implementations vary from traditional IS projects in terms of requirements as; multi-disciplinary teams, agile development with frequent business user check-points, data profiling, visualization, non-deterministic outcomes, change management, optimizing resource management.

There has been little research conducted related to IT professionals and big data. Specifically, to our knowledge, there have been few studies to determine critical success factors of Big Data projects and examine the relationship among these factors. This research can help organizations in general to identify factors that impact success – as perceived by practitioners and professionals – on Big Data projects.

Background of the Study

There is a long history of data analysis and the application of Big Data within an organizational context. The first large-scale methods for metadata creation and analysis (an arrangement of clay tablets revealing data about livestock) have been linked to the Sumerian people, lived in the early Bronze Age (Erikson, 1950). Similarly, card catalogs, and other information methods used in libraries (Lee, Clarke and Perti, 2015), are forerunners to the large-scale digitized metadata collections of today, as they too were technologies used for gathering and storing facts about data in comprehensive and systematized ways (Lee, Clarke and Perti, 2015). The rise in digital technology is leading to the overflow of data (Gog et al., 2015), which constantly requires more updated and faster data storage systems (Sookhak, Gani, Khan, and Buyya, 2017). The recognition of data excess started as early as the 1930s but was not actually named Big Data until the mid-1990s by John Mashey (Kitchin and McArdle, 2016). Now, Big Data has many applications in a wide variety of fields and problem domains.

There has been significant growth in the use and application of Big Data technologies (Sim, 2014). This growth has been fueled by several factors. First, the amount of data that organizations collect and store is difficult to manage because of the variety, velocity, volume, and veracity of the data. Big data refers to the scale, speed, certainty, and diversity of the data (M. Chen, Mao, and Liu, 2014). Thus, there is a need for solutions that assist in solving data-related problems (H. Chen, Chiang, and Storey, 2012). It is necessary for organizations to be able to turn data into useful and actionable information. Second, companies are facing pressure to leverage their data in order to reduce costs and remain competitive (Larose and Larose, 2014).

There are significant benefits of implementing Big Data projects within organizations. In a case study related to Big Data, the outputs of the implementation were able to help reduce fuel costs, predict the overall health of a vehicle, and optimize driver behavior by quantifying the effect on vehicle performance (Melli et al., 2012). Organizations can realize benefits such as improved product safety and product usability as well as process optimization in advanced manufacturing (Zheng et al., 2014). Grocers have used Big Data to optimize the layout of floor space in order to increase profit and enhance customer experiences. Furthermore, Big Data has helped retailers build customer loyalty programs by classifying the most profitable customers (Mittal, 2014). Since Big Data can be used to enhance organizational decision-making and provide organizations with benefits, it is necessary to investigate how to ensure Big Data projects' success.

Organizations face several challenges with respect to implementing Big Data projects. Many organizations experience challenges in terms of completing a successful IS implementation efforts. Innotas, an IT project portfolio management organization, found that more than 50% of businesses surveyed had an IT project fail during 2013 (Florentine, 2013). Altuwaijri and Khorsheed (2012) reported that 44% of all IS projects are partial failures. Projects that are partial failures perhaps did not finish on time or within an allocated budget. Furthermore, this research team reported that 24% of all IS projects end up as total failures. Total failures resulted in IT implementations that were never completed or the resulting system was never adopted and used.

The primary reason for so many failures was due to a lack of resources to meet project demands. In addition to this reason, other reasons for project failures include poor planning, lack of clearly defined problems, lack of top management support, poor project management practices, misunderstanding user requirements, lack of end-user involvement, changing scope and objectives, insufficient or inappropriate staffing, and lack of team knowledge and skills (Al-Ahmad et al., 2009; Kerzner, 2014). The Project Management Institute (PMI) also conducted a study about how top management and executive sponsors support a project. They reported that being an executive sponsor of a project would require balancing both trust and involvement with a project team (PMI, 2015).

Resource requirements and high implementation costs have been blamed for the significant failure rates of Big Data projects. One research team described how the costs involved in a Big Data project are difficult to estimate (Marban, Menasalvas, and Fernandez-Baizan, 2008) and there are many different types of costs that are incurred throughout a Big Data project (Tabladillo, 2009). Costs involved in a Big Data project may include software licensing or purchasing fees, hardware or maintenance, data collection, data preparation, and staff professional development. There are also qualitative costs such as organizational culture changes associated with technology implementations (Tabladillo, 2009).

Purpose of the Study

It is very clear from literature and feedback from the practitioners of the field that Big Data is here to play a role in our future (Gamage 2014; Burg 2014; Allouche 2014; Halaweh and Massry 2015; Wamba et al., 2015; Wixom et al., 2014; Xu et al 2015; Chen et al 2012; Forrester 2012).

This study focuses on identifying the key areas – also called "Critical Success Factors" (CSFs) – essential for achieving success in Big Data projects. The main purpose of this research is to build on the current diverse literature around Big Data by contributing discussion and data that allow common agreement on factors that influence successful Big Data projects. The research also validates the CSF scale and relational CSF model statistically. While individual and technical factors have been explored as they relate to Big Data success, there is a gap in the literature in determining the critical factors in the light of the views of Big Data experts. Even though critical success factors have been discussed previously as being related to IS success, it has not been associated with Big Data project success. This study focuses on three significant drivers. First, Big Data projects have been described as complex and costly endeavors (Akkaya and Uzar, 2011; Delen, 2015). Second, Big Data projects require a strong understanding of both the problem domain and skills in knowing managerial requirements, which Big Data projects can be used for a given problem. Third, IT projects, and business intelligence-related projects, in particular, have high failure rates. Gartner Research found that at least 30%

of Big Data projects did not meet business needs and project objectives (Saran, 2012). Huang et al. (2012) and Sim (2014) indicated that research is needed to investigate factors that relate to Big Data, data analytics, and business intelligence implementation success. The research questions being investigated are based on Big Data project success. The researcher explores the following questions with this research; *"What are the CSFs that impact perceived project success in Big Data projects?"* (qualitative research question) and *"What are the relationships among the CSFs?"* (quantitative research question). In this context, CSFs and several hypotheses mentioning relations between the CSFs will be examined. The relations among CSFs will be visualized and tested in a relational model via Structural Equation Modeling (SEM).

This research may contribute to the IS success and Big Data literature by determining which success factors are critical for projects and examine whether there is a statistical relationship between the factors. It is expected that most of the critical success factors will be consistent with previous studies within the IS success literature (Almajed and Mayhew, 2014; Palanisamy et al., 2010); but also there will be Big Data specific factors. Enlightening predictor success factors of Big Data projects are crucial since researchers have stated the need for determining the factors associated with Big Data (Sim, 2014).

Significance of the Study

In the 1960s, the concept of Critical Success Factors was introduced and can be defined as elements essential to execute the project successfully. The literature suggests that CSFs are important factors for IS projects (Abdekhoda et al., 2015; Almajed and Mayhew, 2014; Liu, Wang, and Chua, 2015). Many studies, as discussed in the literature review, throw light on various critical success factors identified and validated for IS projects. Scholarly articles have investigated if individual, technical, and some organizational factors are related to IS success (Ang, 2009; Bole et al., 2015). These CSFs have been categorized so far into generic groups such as People, Process, Technology, etc. Categorization of these CSFs for Big Data projects is a gap that needs to be filled.

This study contributes to the existing literature that pertains to project success by determining the critical success factors for Big Data projects and validating a theoretical research model. Critical success factors have been studied extensively as it relates to IS success (Haque and Anwar, 2012; Maclennan and Van Belle, 2014). But current research is inadequate to enlighten specifically Big Data projects. These projects require knowledge and managing skills in technical, managerial, and analytical perspectives (Jin et al., 2015; Villars et al., 2011). Big data projects are IS projects in basis but differs in data quantity so in capturing, storing and analyzing; this brings several advancements and challenges within (Saltz, 2015). To deploy and exploit Big Data in an optimal manner, it is necessary for the organization to pay more attention in managing these projects more efficiently.

Currently, Big Data research is concentrated on enhancing data models and algorithms; however, the best approach to execute these projects must also be studied. Further complicating the situation, Big Data projects are exploratory in most cases, and accordingly, the projects lack clear business requirements with subsequent results and they are not easily validated (Saltz and Shamshurin, 2016). Moreover, teams performing data analysis and data science work operate in an ad hoc fashion, where a trial and error process is used to identify the right tools and accordingly involves a low level of process maturity (Saltz and Shamshurin, 2016). The results of this research would shed valuable insights regarding Big Data project success.

Prior to this current study, the CSFs play a crucial role for successful completion of Big Data projects. The projects were slightly examined and the relationships between the CSFs were unknown as they are never been statistically tested and validated. The most complete information, regarding the CSFs for a Big Data projects, can be received from Big Data professionals within those departments for Big Data projects (Sivarajah et al., 2017). Accordingly, this study is conducted with 17 Big Data experts and 827 Big Data professionals. At the end of the study, five CSFs emerged and a statistically reliable and valid scale and a relational research model are added to the literature and further tested empirically.

This study can be evaluated as significant for both academic and practical perspectives. In terms of academic contribution, our original research goal is to close the gap in the literature, regarding Big Data project success. Relational representation of critical success factors in a statistical model and developing a CSF scale is a new approach for both Big Data and critical success factors literature. The methodology of this study strengthens the findings. Several semi-structured interviews and a two-round Delphi study are conducted to enlighten the critical success factors. The predicted relationships among the factors are visualized on the research model and subsequently, quantitative data is gathered from 827 Big data professionals in order to validate the theoretical research model statistically. The results could extend the IS success model by introducing critical success constructs in Big Data implementations. The IS success model (DeLone and McLean, 1992; 2003) includes concepts of information and data quality, service quality, and system quality. Its weakness is that it neglects organizational factors such as management support, team or project related issues. The study presents practical contributions for Big Data project owners. The practical usage of this study can help organizations to identify factors contributing to the success or failure of Big Data projects. The research is based on the knowledge and experience of a great group of Big Data experts and workers. Statistically significant results and validated relations between the CSFs promises to take smarter steps while planning a Big Data project. According to a report from Gartner (2017), 60% of the Big Data projects end with disappointment. Some experts claim that reality is worse and 85% of the Big Data projects fail (techrepublic, 2017). This picture gives us an opinion about the challenges the industry faces in order to reach success. The study could contribute to professional practice by assisting top management teams with identifying possible problem areas when implementing Big Data projects. In addition, this study could provide Big Data professionals with an increased understanding of how Big Data projects are impacted by the presence or absence of suggested CSFs. Big Data professionals may also be prepared to explain how CSFs are necessary for successful project completion. The study is part of the broader field of Big Data and business intelligence initiatives, where organizations use these technologies as part of their data and information management strategy to achieve enhanced decision-making capabilities. The thesis contributed to the body of literature by describing which organizational factors are related to Big Data project success and investigated the existing proposed relationships. Thus, this research uncovers the CSFs of Big Data projects, which we hope to help Big Data project owners to create a better project plan with more chances to meet the expectations.

Methodology of the Study

Given the above and the early nature of this concept, the main purpose of this research is to investigate CSF in Big Data projects, driven primarily around premise of the Big Data experts' arguments regarding Big Data projects. As a field of study, this thesis could lead innovations and strategic results (Galbraith 2014, Church and Dutta 2013, Brynjolfsson and McAfee 2013, Wamba et al., 2015, Halaweh and Massry 2015). There have been calls suggesting that CSF is strategic (Jelinek and Litterer 1988; Head 2009) and experience with many methods that tie in with Big Data historically as evidenced by Weisbord (2012) analysis of CSF history.

Unfortunately, very limited amount of existing data, framework and variables exist concerning successful Big Data projects. It was, therefore, important to formulate methods that would allow us to collect data, review, analyze, deduce a model, formulate a theory and finally test the phenomenon statistically.

The best approach for such a study was mixed methods utilizing Constructivist Grounded Theory. Mixed methods allow for the integration of qualitative and quantitative data within a study to provide a more complete analysis of the research problem being investigated (Creswell and Plano Clark, 2011). It allows, especially for an early concept, data to be built and further explored using a secondary method. Grounded theory allows the researcher to begin with the question, collect data, examine ideas and concepts, extract and categorize, use data, and form the basis of a new theory. This new theory can then be applied and tested statistically. To successfully accomplish this, the approach of this thesis was fragmented into a three-part mixed methods study.

A qualitative section utilizing semi-structured interviews and Delphi study with experts in the field followed by a quantitative section to test relationships between core concepts derived from the qualitative section. First, conducting a qualitative study is suitable for the current research, since the study is designed to investigate perceptions, experiences, and ideas (Ashby, Fryirs, and Howitt, 2015; Merriam, 2014). The qualitative technique is also useful for gathering a consensus opinion not found in the literature, an effort that would not be feasible with quantitative or mixed method approaches (Rees, Rapport, and Snooks, 2015). The qualitative portion of the study was done first, which allowed relationships to be tested later in a quantitative manner using statistical techniques. The knowledge gained through such process allowed the quantitative section to be further insightful, concentrated and exploratory in nature.

Furthermore, it is important to note that this research also examines the relations among variables (Kerlinger and Lee, 2000) as it is being conducted to determine the CSF relationships for successful Big Data projects. Standard strength and direction of relationships between variables are examined and predictions provided given the strength and conclusive nature of the variables within the study. The step by step process to investigate the research problem is as follows:

1. The first step was to be formally educated on both of these topics. As the researcher was already on the journey to obtain a Ph.D. in Management Information Systems, it was vital to enhance her knowledge on Big Data. In order to have a professional standpoint, the researcher worked with a consultant on Big Data projects. Even so, education was needed to familiarize with various tools and techniques that professionals use in this trade every day. The researcher started by speaking to multiple global, startup and mid-size organizations, joined related LinkedIn professional discussions groups and looked up reading the latest on Big Data. All this was done to increase the knowledge and skill level with the goal of being able to conduct semi-structured interviews and Delphi study and have detailed conversations with professionals. This was an evolving process, started in June 2017.
2. The second step was conducting semi-structured interviews with experts about what does “success” mean in Big Data projects. The research model consists of the CSF variables and “success” as the dependent variable. Delphi and computer-assisted telephone interviewing (CATI) rounds are utilized to form the CSFs. The semi-structured interviews with experts aim to enlighten what “success” meant for a Big Data project. The analysis of the semi-structured interviews generated the keywords and finally the scale for success variable.
3. The next step was to start conducting Delphi study with professionals who have worked on and implemented Big Data projects, programs, and solutions. This was the qualitative phase. The researcher utilized their personal and professional network to locate professionals and organizations who had implemented Big Data initiatives, solutions, projects and/or programs and who were willing to speak about their experiences. This is commonly referred to as the “purposeful sampling” technique in qualitative research. The purpose of the Delphi study was to get feedback on success factors of Big Data projects. The initial goal was to speak with roughly 10-20 professionals regardless of industry, profession or location as deemed sufficient in Delphi studies.

4. After conducting the Delphi Study, the researcher would look for common success factors that can be grouped into concepts and further into categories of CSF, measured as variables using a survey.

5. The final step was the creation of the survey. This was the quantitative (survey) portion of the mixed methods study. This would then allow the researcher to run statistical procedures to determine various CSFs for Big Data projects.

The integration of the qualitative and quantitative designs for this research allowed the researcher to help better understand, compile and relate Big Data Projects with critical success factors. This integration, as Creswell and Plano (2011) elude allow for a single study to provide a more complete analysis of the research question being explored. In other words, we take one set of data, perform analysis and apply our insights to build the other data set. This helps to further expand on the knowledge gleaned from just the primary method. As such, this two-part design allowed the researcher to holistically look at factors impacting successful Big Data Projects. As we will review here, the qualitative research was conducted prior to the quantitative study. The learning's gathered from the initial qualitative analysis allowed the researcher to create a scale to statistically analyze the hypotheses and the quantitative research question.

Assumptions and Limitations

This study included several assumptions regarding data gathering and analysis. We assume that the participants answered our questions in the semi-structured interviews, Delphi study and CATI survey honestly. They didn't have any bias in answering reading or listening the questions. They had basic knowledge of the premise of each question as given in the instructions for the question. Since it is a convenience sample, we assume that it is representative of the total population of Big Data professionals.

This study also encompassed the following limitations: The findings are not necessarily generalizable to the entire population of IT experts. The participants were not compensated for their participation in the study.

A few points regarding implications for the study to keep in mind are: (a) there were only a small number of experts who were attended the Delphi study and they were all found via professional and personal contacts of the researcher, (b) the Delphi categorizing and inspection of themes was conducted solely by the researcher and subject to interpretation,

(c) the survey questions were formed by the researcher based mostly from literature review and expert opinions, (d) the survey questions had multiple questions, measuring similar characteristics and that may have distributed the impact of some of the factors and (e) anonymity was a very important factor to the experts. Many didn't want to attend the interviews or Delphi nor did they want to be recorded. The author provided as much leeway as possible in answering questions, opting out of the study and minimizing the use of competitive knowledge.

The following chapters provide the details of this research, discussion, and findings. Chapter 1 and 2 provides a comprehensive review of the literature including search criteria, definitions and gap this research is addressing. The methodologies used to study the research are provided in Chapter 3 with the results and analysis of data presented in Chapter 4 and 5. Chapter 6 discusses those findings as well as implications to theory and practice and presents a summary of the study as well as areas for future research and limitations.

CHAPTER 1: UNDERSTANDING BIG DATA

Big data has been one of the major areas of focus in the field of data management. Big data provides the business solutions which help the organizations making their decisions. Current growing value for the data helps organizations innovate quickly the optimum usage of data and keep up the edge (Lukoinova and Rubin, 2014).

Implementation of methodologies should be in context with a technology base that is growing to be a moving target. The main technology behind fostering the rate of innovation in big data platforms and solutions is the open source technology development and delivery model. Organizations face challenges with evolving business needs and technologies, organizations hold the flexibility for the platforms, solutions, and evolving their capabilities so that they derive value and positive insights from their big data investments (Nimmagadda and Dreher, 2013).

According to the latest Worldwide Semiannual Big Data and Analytics Spending Guide from International Data Corporation (IDC), worldwide revenues for big data and business analytics (BDA) will grow from $130.1 billion in 2016 to more than $203 billion in 2020 (IDC, 2015).

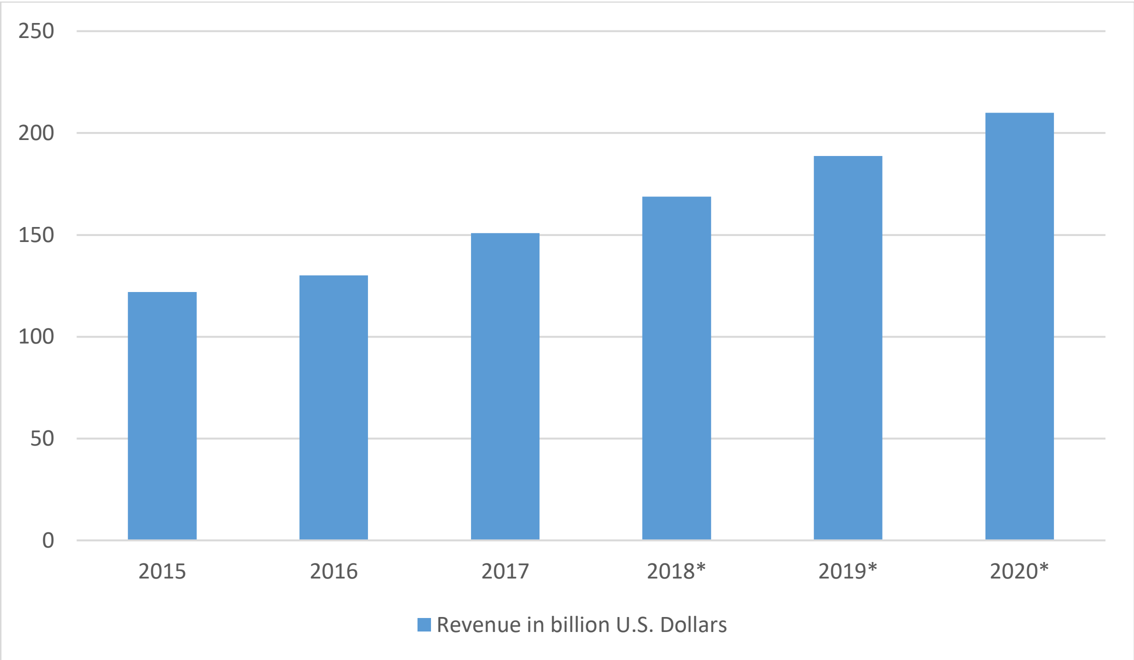

Figure 2: Market Predictions on Big Data (USD Billion)

Source: IDC (2015)

Organizations which handle the big data and implement its methodologies are expected to make 40% more profits than regular software industry does in the current scenario. The increasing value for big data makes it easier to predict the gains for the organization in the future. Organizations currently lack the human resource and talent which can give them the best big data engineering experience and help them grow.

The era of big data has established a new path for exploring data in newer forms and finding different ways to handle the data on a large scale. Although processing and maintaining a large data is a challenge, big data challenges have given the scope to find a solution for these challenges and implement them for a better data environment (Chen et al., 2013). Big data has been into existence since the 1990s and data integration has been one of the major challenges since then. Data Integration in large: Challenges of Reuse, a research paper which was published in 1994 signifies the existence of big data from 1990‟s.

1.1. An Overview

Evolution of large data sets from major industries is termed as big data in the field of data science. The first large-scale methods for metadata creation and analysis (an arrangement of clay tablets revealing data about livestock) has been linked to the Sumerian people, active in the early Bronze Age (Erikson, 1950). Similarly, card catalogs, and other information methods used in libraries (Lee, Clarke and Perti, 2015), are forerunners to the large-scale digitized metadata collections of today, as they too were technologies used for gathering and storing facts about data in comprehensive and systematized ways (Lee, Clarke and Perti, 2015). The rise in digital technology is leading to the overflow of data (Gog et al., 2015), which constantly requires more updated and faster data storage systems (Sookhak, Gani, Khan, and Buyya, 2017). The recognition of data excess started as early as the 1930s but was not actually named Big Data until the mid-1990s by John Mashey (Kitchin and McArdle, 2016). The sudden increase in the U.S. population, the dispensing of social security numbers, and the wide-ranging increase of knowledge (research) required more detailed and organized record-keeping (Gandomi and Haider, 2015).

Big data can be classified as the large volumes of data-sets with a higher complexity level. Gandomi and Haider (2015), IDC, IBM, Gartner, and many others have contributed with an excellent summary regarding Big Data characteristics. Clearly, size is the first characteristic that comes to mind considering the question “what is big data?” (Gandomi and Haider 2015). Following that, the Three V’s have emerged as a common framework to describe big data (Chen, Chiang, and Storey, 2012; Kwon, Lee, and Shin, 2014): Volume, Variety, and Velocity. There have been more additions: IBM, White (2012) introduced Veracity – the fourth V, SAS introduced Variability and Complexity, the fifth V and Oracle introduced Value as the sixth V. While these are commonly used today there are possibilities with further enhancements more may be added, or defined further contextually. There is even the possibility of having "smarts" added to this volume of data as well. There are questions about the usefulness and life of the data as well.

The concept of big data has been described as "a phenomenon defined by the rapid acceleration in the expanding volume of high velocity, complex, and diverse types of data. Big Data is often defined along three dimensions -- volume, velocity, and variety" (TechAmerica Foundation 2012, p. 7). Many authors will refer to those three characteristics as the 3V's. Others define big data as "datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyze" (MGI, 2012, p. 3).

Despite big data 3V's characteristics - volume, velocity. and variety, some authors write about multiple fourth "V"s such as variability, vulnerability, veracity, and value. The fundamental definition is not affected by many "V"s, but all together they do provide a better understanding of different aspects of big data (Seddon and Curie, 2017). It is anticipated that volume of data will increase 44 times by 2020; velocity will increase as data is brought in from every imaginable device, and variety will increase due to a greater diversity in the data being collected. (Fernandes, O'Connor, and Weaver, 2012).

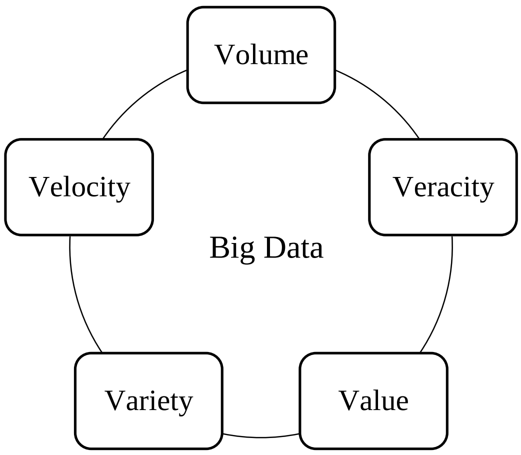

Figure 3: 5V's of Big Data

In order to define Big Data, we look at the definitions for each of the 5 Vs below as they seem to characterize Big Data broadly:

Volume - Volume is the large data-sets that represent big data. Volume makes a huge difference for an organization as the huge data is what they require to make business decisions.

Variety - This represents the different types of data available, such as text, numbers, images, videos, documents, spreadsheets, etc. This signifies the category or type of data something belongs to. The big data comes from different sources which makes it very unpredictable and consists of different forms which are ideally unstructured, structured and semi-structured. The unstructured data has the log files, HTML tags. Structured data consists of the relational database data which is represented in tables. Semi-structured data consists of XML files and data from other text files.

Velocity - Velocity represents the speed of data at which it is transmitted and received from the source and destination. Velocity plays a crucial role in data management as the process flows in the business are highly impacted by the speed of data transfer.

Veracity - Veracity represents the uncertainty of the data as it comes from an untrusted source and needs more optimization. Veracity ideally is characterized by raw data.

Value - Value represents the revenue and market value gained by an organization using the big data. Value is measured in terms of revenue and business's success with their clients using the tools for generating the value for data.

The five v's of big data impact the scope, time, and budget for any project which deals with big data (Yin and Kaynak, 2015). The opportunity cost, ambiguity, and collection ability play a role in authenticity/reliability of the data, the inconsistencies behind gathering and gaining the data and the value derived and implementation costs from the data.

In summary, having gone through the definitions that exist in literature today and having looked at characteristics to date, we are still not close to agreeing on the definition of the term, Big Data. As an MIS (Management Information Systems) scholar, the interest is in all the moving parts that contribute to the definition and success of Big Data. Data is focal to IS tools and methods to derive those valuable recommendations. To take this step, it was first important to search the literature for pre-existing approach and research areas regarding Big data. Literature is examined to reveal the current gap in the field of Big Data. Accordingly, a focus distribution within the field is emerged and presented in Table 1.

Table 1: Literature on Big Data

Issues related to Big Data	References
IT and Big Data Investments	Snow, 1966; MacMillan and Day, 1987; Solow, 1987; Jacobs, 2009; Chen et al., 2012; Forte, 1994; Williams and Williams, 2007; Lee et al., 2014; Powell and Snelman, 2004; Willcocks and Lester, 1996; Willcocks et al., 1999; Brynjolfsson, 1993; Brynjolfsson and Hitt, 1998; Jones et al., 2012; Dos Santos and Sussman, 2000; Lucas, 1999
Basic Research	Gao et al., 2015; Seddon et al., 2010; Chen et al., 2012; Kumar et al., 2013; Goes, 2014; Agarwal and Dhar, 2014; Bharadwaj et al., 2013; Zott and Amit, 2007; Hoy, 2014; Mayer-Schönberger and Cukier, 2014; Vinod, 2013; Rubinstein, 2013; Beyer and Laney, 2012; Dumbill, 2013; Narayanan et al., 2014
Technical perspective	McAfee et al., 2012; Hu et al., 2014; Zikopoulos and Eaton, 2011; Davenport et al., 2012; Boyd and Crawford; 2012; Katal, Wazid and Goudar, 2013; Bryant, Katz and Lazowska, 2008; Madden, 2012; Gandomi and Haider, 2015
Organizational perspective	Lohr, 2012; Bughin, Chui and Manyika, 2010; Marz and Warren, 2015; Mayer-Schönberger and Cukier, 2013; LaValle et al., 2011; Chen, Mao and Liu, 2014; Siemens and Long, 2011; Michael and Miller, 2013; Villars et al., 2011; Bizer et al., 2012
Analysis methods and algorithms	Lazer et al., 2014; Wu et al., 2014; Scott et al., 2016; Rebentrost, Mohseni and Lloyd, 2014
Decision support	Bughin et al., 2010; Schadt et al., 2010; Cole et al., 2012; Brown et al., 2011; Bughin et al., 2011; LaValle et al., 2011; Meijer, 2011; Sobek et al., 2011; Boyd and Crawford, 2012; Allen et al., 2012; Anderson and Blanke, 2012; Ann Keller et al., 2012; Boja et al., 2012; Beath et al., 2012; McAfee and Brynjolfsson, 2012; Davenport et al., 2012; Demirkan and Delen, 2013; Fisher et al., 2012; Gehrke, 2012; Griffin, 2012; Dansion and

Issues related to Big Data	References
	Griffin, 2012; Johnson, 2012; Kolker et al., 2012; Lane, 2012; Ohata and Kumar, 2012; Smith et al., 2012; Soares, 2012; Strawn, 2012; Tankard, 2012; Wagner, 2012; White, 2012
Alternative usage and utilization methods for databases	O'Driscoll, Daugelaite and Sleator, 2013; Demchenko et al., 2013; Madden, 2012
Technical deficiencies and problem-solving	Jagadish et al., 2014; Hashem et al., 2015; Kaisler et al., 2013; Katal, Wazid and Goudar, 2013
Organizational value	Lazer et al., 2014; LaValle vd. 2011; Jagadish et al., 2014
Competitive advantage	Chen et al., 2012; Marz and Warren, 2015; Mayer-Schönberger and Cukier, 2013; LaValle et al., 2011; Chen, Mao and Liu, 2014
Performance improving	Brinkmann et al., 2009; Bughin et al., 2010; Schadt et al., 2010; Brown, et al., 2011; LaValle et al., 2011; Long and Siemens, 2011; Cole et al., 2012; Sobek et al., 2011; Allen et al., 2012; Anderson and Blanke, 2012; Keller et al., 2012; Beath et al., 2012; Boja et al., 2012; Boyd and Crawford, 2012; Chen et al., 2012; Davenport et al., 2012; Demirkan and Delen, 2013; Fisher et al., 2012; Havens et al., 2012; Huwe, 2012; Wagner, 2012; Johnson, 2012a; Soares, 2012; Kolker et al., 2012; Strawn, 2012; Tankard, 2012; White, 2012; McAfee and Brynjolfsson, 2012
Managing with Big Data	George, Haas and Pentland, 2014; Lohr, 2012; Bughin, Chui and Manyika, 2010
New business models, products and services	Bughin et al., 2010; Bughin et al., 2011; LaValle et al., 2011; Brown et al., 2011; Long and Siemens, 2011; Ann Keller et al., 2012; Cole et al., 2012; Beath et al., 2012; Boyd and Crawford, 2012; McAfee and Brynjolfsson, 2012; Davenport et al., 2012;

Issues related to Big Data	References
	Chen et al., 2012; Demirkan and Delen, 2013; Fisher et al., 2012; Gehrke, 2012; Griffin, 2012; Griffin and Danson, 2012; Huwe, 2012; Johnson, 2012; Kolker et al., 2012; Ohata and Kumar, 2012; Soares, 2012; Strawn, 2012; Tankard, 2012; Wagner, 2012
Development of Big Data	Hilbert and Lopez, 2011; Chen, Mao and Liu, 2014; Cukier, 2010; Zikopoulos and Eaton, 2011
Organizational effects	Bharadwaj 2000; Grant 2010; Carr, 2003; Ross et al., 2013; Amit and Schoemaker 1993; Teece, 2014; 2015; Teece et al., 1997; Vera-Baquero et al., 2013; Tonidandel et al., 2015; Kamioka and Tapanainen, 2014; Calvard, 2016; McAfee and Brynjolfsson, 2012; Barney, 1991; Manyika et al., 2011; Knox, 2013; Miller, 2013; George et al., 2014; Davenport, 2014; Mata et al., 1995; Wixom and Watson, 2001; Chae et al., 2014; Chen et al., 2012; Nonaka et al., 2000; House et al., 2002; Dowling, 1993; Lavalle et al., 2011; Grant, 1996; Bhatt and Grover, 2005; Cohen and Levinthal, 1990; Nonaka and Teece, 2001
The potential of Big Data	Wielki, 2013; Linoff and Berry, 2011; Saltz, 2015; Al Nuaimi et al., 2015; Elragal, 2014; Hazen et al., 2014; Simon, 2013; Işık et al., 2013; Dutta and Bose, 2015; Ohlhorst, 2012; Rajpurohit, 2013; Yin and Kaynak; 2015; Franks, 2012; Russom, 2013; Ayankoya et al., 2014
Research by industry	Retail: (Brown et al., 2011; Lee et al., 2013; McAfee and Brynjolfsson, 2012) Healthcare: (Brinkmann et al., 2009; Field et al., 2009; Callebaut, 2012; Chen et al., 2012; Cole et al., 2012) Ecology: (Hochachka et al., 2009) Education: (Long and Siemens, 2011; Soares, 2012) Government: (Sobek et al., 2011; Chen et al., 2012; Mervis, 2012)

Issues related to Big Data	References
	Manufacturing: (Brown et al., 2011) Service: (Acker et al., 2011; Demirkan and Delen, 2013; Johnson, 2012; Kauffman et al., 2012; Kolker et al., 2012; Kubick, 2012; McAfee and Brynjolfsson, 2012) Technology: (Bradbury, 2011; Reddi et al., 2011; Allen et al., 2012; Chen et al., 2012; Burges and Bruns, 2012; Smith et al., 2012) Miscellaneous: (Jacobs, 2009; Bughin et al., 2010; Schadt et al., 2010; Alexander et al., 2011; Brown et al., 2011; Bughin et al., 2011; Kiron and Shockley, 2011; LaValle et al., 2011; Chen et al., 2012; Cole et al., 2012; Davenport et al., 2012; Griffin, 2012; Dansion and Griffin, 2012; Kauffman et al., 2012; Mervis, 2012; Strawn, 2012)

1.2. Current Status

IT departments do not measure the growth of Big Data by the number of records that are in storage but by the amount of space required to store the records (Kitchin, and McArdle, 2016). To illustrate this point Abbasi, Sarker, and Chiang (2016) noted this space now consists of "Gigabytes, Terabytes, Exabytes, and Petabytes" (p. 5) versus previous traditionally records based number approaches to data management. As well as the expanding data size, the monetary value of Big Data also increases with a very high rate. The global big data market size was valued at USD 25.67 billion in 2015 and is expected to witness a significant growth over the forecast period (Grand View Research, 2016).

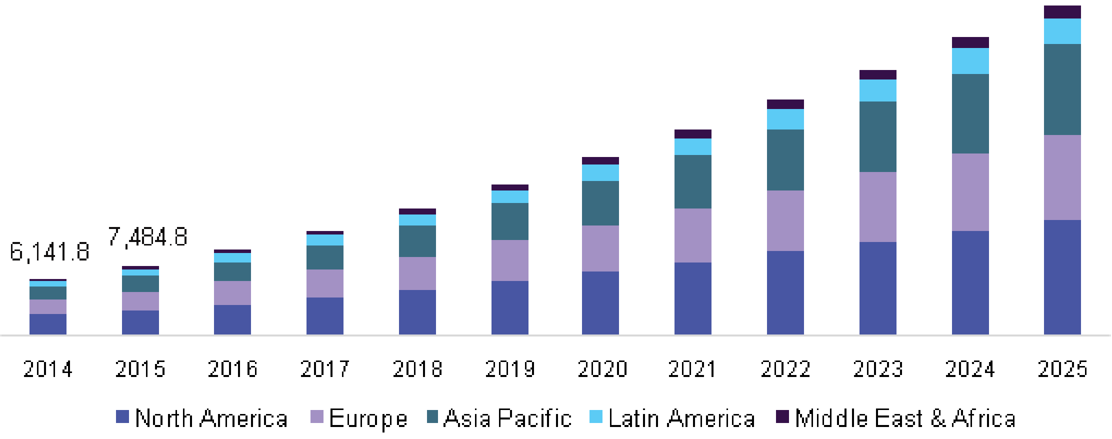

Figure 4: Big Data Market by Service (USD Million)

Source: Grand View Research (2016)

This widespread growth pattern stems from many different sources such as social networking sites, wired and wireless broadband access, and the widespread use of search engine sites (Hashem, et al., 2015). Additional non-interactive devices are also filling storage such as radio frequency identification (RFID) and sensors associated with the Internet of Things (IoT) (Reimsbach-Kounatze, 2015). Expanding market also affects software markets. Big Data market is shared by analytics, database, visualization and distribution tools, and software. Database related products are the main driver of the commercial transactions.

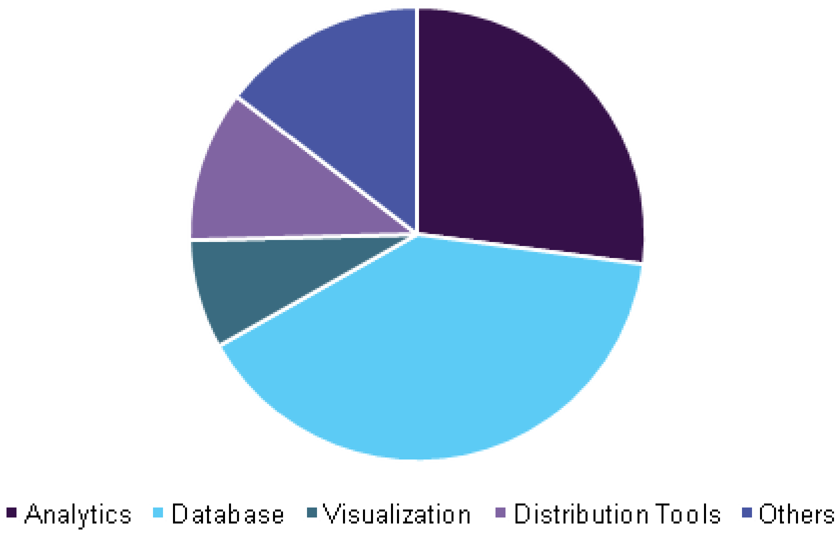

Figure 5: Big Data Market Share by Software (USD Million)

Source: Grand View Research (2016)

This growth is so rapid that both practitioners and academics are trying to keep up with newer and faster analytics and statistics (Gandomi and Haider, 2015). Since the majority of Big Data is unstructured (approximately 95%), the data is harder to process (Gandomi and Haider, 2015). While IoT and Big Data analytics appear to have incredible potential for converting various businesses, many academics and industry experts are struggling to comprehend these ideas and capture the business value in joining IOT and big data analytics (Riggins and Wamba, 2015). In addition, very few academic studies exist assessing the real potential of IoT and Big Data analytics (Riggins and Wamba, 2015). However, as noted previously, there are many different projects underway to resolve the difficulty of converting raw data into information and then into knowledge.

In recent years, the explosive growth of data has been observed in numerous industries like e-commerce, health, social networks, etc. Access to preferred data in such massive datasets necessitates sophisticated and effective gathering methods. In the past, special algorithms have served as common descriptors for numerous tasks including image cataloging and recovery (Ahmad et al., 2018). The algorithms perform extremely well when equated to hand-crafted queries and filters. However, these algorithms are typically high dimensional, necessitating a lot of memory and CPU for indexing and gathering. For extremely large datasets, use of these high dimensional algorithms in raw usage becomes infeasible (Ahmad et al., 2018).

Boone, Skipper, and Hazen (2017) conducted a study to Increase request for receptive, cost effectual, and maintainable procedures that necessitated service parts deliberation and acceptance of new industry models and resolutions that cover the complete life-cycle of merchandises. Many organizations are looking to big data information. Though service parts supervisors have long trusted examination and optimization, big data information is thought to be more incorporating and thus particularly capable. Hereafter, big data and its associated uses are suggested as a means of refining service parts supervision practices. More precisely, information gathered from consultations with service parts supervisors is used to build a basis describing the encounters of service parts supervision. This background then aids as the foundation for big data connected suggestions for overcoming the emphasized encounters. Thus, the examination answers the demand for service parts supervision connected backgrounds while developing a starting point for suggestions for supervisorial thought and intellectual examination (Boone et al., 2017).

The advance of Big Data that of personal data, in particular, dispersed in numerous data sources presents huge opportunities and understandings for companies to discover and influence the importance of linked and assimilated data. However, privacy fears impede distribution or trading data for connection across diverse organizations (Vatsalan et al., 2017). Privacy-preserving record linkage (PPRL) purposes to address this situation by recognizing and joining records that match to the identical real-world individual across numerous data sources stored by diverse parties without revealing any sensitive data about these individuals (Boyd et al., 2017). PPRL is progressively being required in numerous practical application areas. Instances include public health, observation of crime, deception exposure, and national security. Big Data and PPRL creates numerous challenges, with the three main ones being (a) scalability to several large data warehouses, due to their considerable volume and the movement of data within Big Data solutions, (b) attaining high quality and effects of the link in the occurrence of variety and veracity of Big Data, and (c) maintaining discretion and confidentiality of the individuals represented in Big Data pools (Vatsalan et al., (2017).

IT departments have encountered several new skills due to Big Data, including how to gather, allocate, accumulate, clean, examine, filter, examine, portion, protect, and envision data (Purdam, 2016). Considering the problem of accumulating and saving big data, an array of new systems emerged in recent years to handle these kinds of big data encounters (Gudivada, BaezaYates, and Raghavan, 2015). Big Data computational analysis is considered an important aspect to be further enhanced to intensify the operational margin of both public and private initiatives and signifies the next frontier for their modernization, competition, and throughput (Esposito et al., 2015). Big Data is typically formed in different sectors of private and public organizations, often physically distributed throughout the world, and are categorized by a large size and variety (Rajan, 2015). Therefore, there is a solid need for strategies handling larger and faster amounts of data in settings characterized by multifaceted event processing programs and multiple mixed sources, dealing with the numerous issues related to resourcefully gathering, examining, and distributing them in a fully controlled manner (Esposito et al., 2015). This necessity leads to newer, faster, and better software (Kim et al., 2016).

Developing and running software creates large quantities of raw data about the development process and the end user usage. This information can be turned into creative perception with the assistance of skilled data scientists (Kim et al., 2016). Unfortunately, data scientists with the skills to analyze these very large data-sets are difficult to come by (Hilbert, 2016). Data scientist comes in many different forms such as Insight Providers, Modeling Specialists, Platform Builders, Polymaths, and Team Leaders (Kim et al., 2016).

Many “big data” and “fast data” analysis methods such as Hadoop, Spark, and Storm have come from the Apache foundation (Dimopoulos, Krintz, and Wolski, 2016). These programs are used by analysts to implement a variety of applications for query support, data mining, machine learning, real-time stream analysis, statistical analysis, and image processing (Dimopoulos, Krintz, and Wolski, 2016). Another software platform named R is a free, prevailing, open source software platform with widespread statistical computing and graphics abilities (Xu, et al., 2016). Due to its advanced expressiveness and many domain explicit packages, R has become the ‘lingua franca’ for many parts of data analysis, acquiring power from community-developed packages (Xu et al., 2016).

With the extremely rapid growth of information and intricacy of systems; artificial intelligence, rapid machine learning, and computational intelligence methods are highly required. Many predictable computational intelligence methods face constraints in learning such as intensive human involvement in addition to connection time. However, effective learning algorithms offer different yet significant benefits including rapid learning, ease of execution, and minimal human involvement. The need for competent and fast execution of machine learning methods in big data and dynamic changing methods poses many research encounters (Sun et al., 2017). Big Data, from an industry point of view, is leading the way to newer and better methods of doing business (Shin and Choi, 2015). The common view, in most industries, is that growth of Big Data, though difficult now, will achieve a status that is manageable and thus controllable in the future (Kitchin and Lauriault, 2015). German officials and scientist believe this so strongly that they are referring to Big Data as the Fourth Industrial Revolution and calling it “Industrie 4.0” (Yin and Kaynak, 2015). Akter and Wamba (2016) used the phrase “the next frontier for innovation, competition, and productivity.” Industries such as retail, are utilizing Radiofrequency Identification (RFID) marked merchandises to develop marketing campaigns based on the movement of merchandise (Cao, Chychyla, and Stewart, 2015).

1.3. Organizational Effects

Akter and Wamba (2016) stated the definition of Big Data is more than merely larger storage or the gathering of data from social media sites with millions of members. Bigness is an indication of scalability issues in one or more extents — the four Vs of variety, velocity, veracity, and volume (Abbasi, Sarker, and Chiang, 2016). Big data is an inaccuracy, suggesting that bigness is a fundamental characteristic of a dataset. Rather, Big Data defines the association between a dataset and its usage framework (Akter and Wamba, 2016). A dataset is too large for a specific use when it is computationally not feasible to convert the data using traditional or outdated software tools (George et al., 2016). With the immense amounts of data currently available, businesses in nearly every industry are focusing on manipulating data for the competitive benefit (He et al., 2015). A key challenge for IT researchers and IT experts alike is that data growth rate is exceeding the ability to maintain the required hardware and necessary software to manage the high volume of data (Saltz, 2015). Simply stating, analyzing "data in motion" creates new encounters because the anticipated patterns and perceptions are moving targets, and this is not the situation for static data (Abbasi, Sarker, and Chiang, 2016). Junque de Fortuny et al. (2013) noted the growth rate is approximately 50% annually or doubling every two years.

Big data provides the business solutions which help the organizations make their decisions. Current growing value for the data helps organizations innovate quickly the optimum usage of data and keep up the edge (Lukoianova and Rubin, 2014). Implementation of methodologies should be in context with a technology base that is growing to be a moving target. The main technology behind fostering the rate of innovation in big data platforms and solutions is the open source technology development and delivery model. Organizations face challenges with evolving business needs and technologies, organizations hold the flexibility for the platforms, solutions, and evolving their capabilities so that they derive value and positive insights from their big data investments (Nimmagadda and Dreher, 2013).

Gobble (2013) and Manyika et al., (2011) identify big data as the next big thing in innovation and the next frontier for innovation, competition, and productivity, respectively. Strawn (2012) called it the fourth paradigm of science. Furthermore, McAfee and Brynjolfsson appropriately categorized their article on Big Data as a management revolution similar to what Ann Keller et al., (2012) termed Big Data as bringing a revolution in science and technology.

The emergence of new technologies, new processes, threats, regulations and thought leadership all affect the organization more than ever. Organizations which handle the big data and implement its methodologies are expected to make 40% more profits than regular software industry in the current scenario. The increasing value for big data makes it easier to predict the gains for the organization in the future. Organizations currently lack the human resource and talent which can give them the best big data engineering experience and help them grow.

Technology is now being established that is able to process enormous amounts of organized and unorganized data from various causes and sources. This information is often denoted to as big data, and opens new areas of study and uses that will have a growing impact in all parts of society (Marvin et al., 2017). Big Data and its velocity are being applied in the food safety area and acknowledged several encouraging trends particularly the speed by which the information is being transmitted. In numerous parts of the world, governments encourage the publication on the Internet of all information produced in publicly financed research projects. This program opens new chances for interested parties dealing with food safety to report issues which were not conceivable before. The use of mobile phones as exposure devices for food safety and the communication of social media as early caution of food safety situations are a few instances of the new improvements that are conceivable due to Big Data (Marvin et al., 2017).

Big data will also offer new potentials for research by allowing access to linked data, medical information, and social media. The total extent of information, however, does not remove and may even intensify systematic inaccuracy. Therefore, procedures addressing systematic error, scientific knowledge, and underlying theories are more significant than ever to confirm that the indicator is apparent behind the noise (Ehrenstein et al., 2017).

The era of big data has established a new path for exploring data in newer forms and finding different ways to handle the data on a large scale. Although processing and maintaining a large data is a challenge, big data challenges have given the scope to find a solution for these challenges and implement them for a better data environment (Du, 2013). Big data has been into existence since the 1990s and ways to success have been one of the major mysteries since then. Data Integration in large: Challenges of Reuse, a research paper which was published in 1994 signifies the existence of big data challenges from the 1990s.

1.3.1. Technology

Netflix analyzes millions of real-time data points that its viewers create, thus helping the firm determine if a pilot will become a successful show (Xu et al., 2015). Facebook hosts over 500 terabytes of data every day – including uploaded photos, likes and users' posts (Provost and Fawcett, 2013). Google alone contributed roughly $54 billion to the US economy in 2009 (Labrinidis and Jagadish, 2012). Akamai Technologies Inc, a leading global Content Delivery Network provider collects and analyzes petabytes of data every day to help its customer base with cloud performance and security initiatives. Amazon, another e-commerce/technology company, utilizes its various data points to ensure personalized experiences for its client base.

Machine learning (ML) is constantly releasing its influence in a wide collection of applications. It has been pressed to the front in current years somewhat owing to the arrival of Big Data and its velocity. ML procedures have never been better guaranteed while tested by Big Data. Big Data empowers ML procedures to expose more fine-grained configurations and make more opportune and precise forecasts than ever before; yet it creates major tests to ML such as model scalability and distributed computing (Zhou et al., 2017). The framework of Big Data is balanced on ML which follows the stages of preprocessing, knowledge, and assessment. In addition, the framework is also comprised of four other components, namely big data, consumer, realm, and method (Zhou et al., 2017).

1.3.2. Healthcare

Burg (2014) argued that Big Data can enable a better and transparent healthcare system. Allouche (2014) identified cost saving and unnecessary procedure reducing capabilities from Big Data. Tormay (2015), identifies pharmaceutical RandD as the engine that fuels the pharmaceutical industry. He claims this engine has been declining in productivity over the last 20 years with increasing costs, demands for better standard care, and concomitant productivity challenges. He believes that data, specifically the fast and voluminous nature along with technological advances will help revitalize this engine. Furthermore, Groves et al. (2013) document the innovations identified because of Big Data projects. Another organization, Intel, announced its Collaborative Cancel Cloud in August of 2015 to enable diagnosing of cancer patients based on their specific genome sequencing and tailor a precision treatment plan for them all based on the concept of Big Data.

1.3.3. Education

Erwin (2015) insists students to be more literate in their abilities to use data. He argues that there is a growing call for students to develop data literacy. His theory is more of a project-based learning where students solve real-world problems with data that is provided to them will enable them to build skills and be able to meet the current demands of business. Similarly, Rijmenam (2014) reasoned changes in the education systems by using Big Data to change the way that students and teachers interact. A more practical example, Gwinnett, in suburban Atlanta, Georgia, is the 14th largest school system in the United States, has 23,000 employees and transports more people every school day than the locally based carrier, Delta Air Lines. All that activity generates information, more and more of it captured digitally and in 2002, as the school system's leaders continued seeking fresh educational solutions, they began to explore how analytics could help how all that information could be investigated for patterns, relationships, dependencies, and predictors.

1.3.4. Public Sector (Government)

Gamage (2016), in his article, examines the opportunities presented by effectively harnessing big data in the public-sector context. He talks about the impact of Big Data and how it will play an important role in the future. Furthermore, he also outlines key challenges to be addressed to adapt and realize the benefits of Big Data in the public sector. Similarly, another article, stemming from SAP’s partnership within the Middle East governments, documents high level of Big Data production, consumption and the need to train public sector to be successful at these opportunities.

1.3.5. Miscellaneous

The Big Data Strategy framework in Servitization as proposed by Opresnik and Taisch (2014) is focusing on new revenue streams and decreasing product-service costs in manufacturing. An optimization model for green supply chain management based on Big Data proposed by Zhao et al. (2017) is a scheme that minimizes the inherent risk of hazardous materials, associated carbon emissions and economic cost. The Cebr (Center of Economics and Business Research) (2012) has anticipated that the benefits of big data innovation opportunities would contribute £24 billion to the UK economy between 2012 and 2017. These opportunities are described to be identifying hidden patterns, better decision making, improving business processes and developing new business models (Halaweh and Massry 2015). There are many more examples of such initiatives and values across industries that organizations are realizing and going back to Gary Kings' quote, at the start of the paper, the accumulation of data is reaching out to every industry and organization across geographies.

In summary, we clearly observe the various opportunities being explored, examined and extracted for the betterment and effectiveness of organizations across the different industries that have successful Big Data projects. This is the very objective behind the reason this research. The research favors utilizing Big Data characteristics and engaging CSFs to lead successful Big Data projects. The research employs semi-structured interviews, Delphi outputs, and empirical data to test this opinion. Alternatively, this will also afford us the opportunity to look at relationships that can impact successful Big Data projects.

1.4. Implementation Challenges

The growth in big data comes along with unstructured data which dominates the data mainly. Therefore, organizations tend to find new methods for handling the unstructured data in large volumes. Organizations implement technologies like Hadoop to handle this data, currently as Hadoop is the only big data tool currently in the market. Hadoop is an open source currently being provided by two organizations Hortonworks and Cloudera. However, some of the tools in Hadoop are not available in the services provided by Cloudera. Big data tools are used individually for different tasks based on the requirement. Some of the big data tools are MapReduce, Yarn, Pig, Hive (Seay et al., 2014).

Another important aspect of big data is the integration services and the storage. Big data integration services include the tools which are used to integrate the data from different sources to gain meaningful insights. Big data integration plays a major role as it provides the organization with necessary data which stands as a base to make the business decisions. KARMA and Talend stand as the best data integration tools currently in the market and Talend has been used by Groupon, one of the largest e-commerce organizations. The different data jobs from different sources were implemented using Talend which gave them good profits on a whole.

The storage of big data has been another important aspect, where the data is generated every second. A load of huge data chunks has to be managed efficiently in order to make use of this data. The storage servers and third-party storage providers bridge the gap between big data and storage. But securing this data is much important for organizations. Accordingly, they tend to use different security measures to prevent the data breach (Ives et al., 1999).

Menon and Hegde (2015) wrote that the indication of the growth of knowledge as an approaching storage and retrieval problem came in 1944 when Fremont Rider, a Wesleyan University librarian, estimated that university libraries in America were doubling in size every sixteen years. Given this growth rate, Rider estimated that in 2040 the Yale Library would contain approximately 200,000,000 volumes, stretching over 6,000 miles of shelves and requiring a cataloging staff of over six thousand people to maintain. While the development of knowledge was generally considered good for humanity, it was leading to a major storage and retrieval situation for libraries (Menon and Hegde, 2015). As the amount of data continued to multiply in the ensuing decades, organizations began to design, comprehend, and execute centralized computing methods that would allow them to automate their inventory systems (Erikson, 1950). As these methods began to mature across organizations and develop within enterprises, organizations began to apply the analysis of the data to avail themselves with solutions and insight that would allow them to make improved business judgments i.e., business intelligence (Wang, 2016).

Wixom et al. (2014) believed that with business intelligence continuing to grow rapidly, the challenge of management and storage quickly became a real issue within IT departments. To offer more functionality, digital storage had to become more cost-effective (Chang, and Wills, 2016). This challenge led to the advent of Business Intelligence (BI) platforms (Wang, 2016). As these BI platforms continue to develop, the data gathered enabled and will enable companies, scientific and medical researchers, our national defense and intelligence organizations, and other organizations to create innovative breakthroughs (Wixom et al., 2014).

At the same time as it problematized data overflow, the Big Data industry was also involved in spreading the myths such as, methodological issues no longer mattered, Big Data provided a comprehensive and unbiased source of data on which to base decisions about data, and in hyping the promises of Big Data sets (Kimble and Milolidakis, 2015). Like Wes Nichols (2013), co-founder and CEO of MarketShare, a predictive-analytics company based in Los Angeles, many market researchers placed a great amount of confidence in the usability of large datasets that concurrently produced and analyzed data. In the past, associating sales data with a few dozen isolated advertising variables used to be acceptable (Oh and Min, 2015). However, many of the world's biggest companies are now deploying analytics 2.0 (Jobs, Aukers, and Gilfoil, 2015), a set of practices that compute through terabytes of data and hundreds of variables in real time to define how well advertising touch points interact (Fesenmaier et al., 2016). This resulted in 10 to 30 percent improvements in marketing performance.

Instead of data storage and integration issues, big data visualization has also been one of the major challenges for organizations. Therefore, they tend to build data samples to build on the tools for data visualization. Visualizations help to understand the data better than its original form. Visual analytics can be integrated with big data to understand and implement the data for a better business.

Some of the data visualization challenges identified is given by:

Meeting the need for speed: Many organizations use data for their business growth. And this data comes from numerous sources. Organizations tend to use different methods to keep this data organized to derive the insights. Few of these methods are discussed below:

- Visualization of data which helps them to perform the analysis more efficiently.

- As the challenge grows with more granularity constraints, some of the organizations are using parallel processing hardware units to grab the large crunch of data at once. This helps them to perform data optimization quickly.
- Grid computing has been another approach to grab a large amount of data in smaller units of time.

Understanding the data: Getting a meaningful insight from the data is a huge task as the data comes from different sources. For an instance, data coming from a social networking site, where it is important to know the user from whom the data is gathered from. Implementation of visual analytics without the proper context can prove useless for organizations. Expertise at domain level is very much needed in this case as the analytics team needs to be aware of the data sources and the understanding of consumers' data usage and interpretation.

Addressing data quality: Although the data is collected and processed quickly to serve the purpose, it is very likely that a data which has no context is of no value to the consumer. To address this challenge, organizations need to have an information management team to analyze the data and assure that the data is clean. Visualizing this analyzed data can be of a huge value and source of information for organizations and the consumers.

Displaying meaningful results: Graphical representation of huge data is a common challenge faced by organizations. For an instance, consider the data from a huge retail business. When the stock-keeping unit (SKU) data which has more than a billion plots, it's difficult to have a look at each plot and speculate. Therefore, data needs to be split into clusters and the small data is to be separated from the big ones to derive the insights.

Dealing with outliers: Big data can be represented in the form of tables and sets as well, but this would be a challenge to viewers. The efficient way of displaying the big data would be by implementing the visual analytics and represent the data in the form of graphs and charts. This would make it easier for the viewers to understand and spot the growth.

On a whole, data visualization plays a key role in big data as this is the main source of information for organizations. The simpler the form of representation is, the simpler would be the analysis and derivations from this data.

Organizations face many challenges with big data in terms of storage, visualization or integration, therefore they look for new solutions and tools which can handle the big data efficiently. This clearly states that big data is the new challenge which needs more research for its development.

1.5. Big Data Projects

In today's world, it is necessary to use the data or information available in a wise manner to make effective business decisions and define better objectives (Laudon and Laudon, 2016). If the information available is not utilized to its full extent, organizations might lose their reputation and position in this competitive world. However, data needs to be processed appropriately to gain constructive insights from it, and the heterogeneous nature of this data makes this increasingly more complex and time-consuming (Ang and Teo, 2000). The ever-increasing growth of data generated is far more than human processing capabilities and thus computing methods need to be automated to scale effectively (Das et al., 2015).

Variety, velocity, and volume were identified as three key attributes of Big Data by Laney in 2001. Many authors and business specialists modify these attributes. In 2012, IBM added a fourth dimension to this called veracity. Ebner et al. (2014) define Big Data as "a phenomenon characterized by an ongoing increase in volume, variety, velocity, and veracity of data that requires advanced techniques and technologies to capture, store, distribute, manage, and analyze these data." Big Data helps to unlock potentials of different fields like predictive modeling, data integration, network analysis, natural language processing, etc. Thus, Big Data technologies have huge economic potential that should be harnessed by executives in a proper manner. Questions related to the IT infrastructure, capturing crucial information, analytical requirements, etc. should be asked by the executives to determine the way in which way Big Data solution needs to be handled.

In recent times, the research on Big Data has been always concentrated toward creating better algorithms and designing robust data models (Saltz and Shamshurin, 2016). However, not much work has been done regarding finding out the best methodology for executing such projects (Ahangama and Poo, 2015) (Saltz, 2015). The exploratory nature of Big Data projects demands a more specific methodology that can handle the uncertain business requirements of such projects (Saltz, 2015). According to a survey carried out by Kelly and Kaskade (2013), "300 companies reported that 55% of Big Data projects don’t get completed and others fall short of their objectives.” The reasons for such project failure can be identified at the beginning of the project or can be reduced at a later stage by some coordination methodology (Saltz, 2015).

A well-defined Big Data analysis project methodology would help to address different issues like roles and responsibilities of team members, project stakeholders, expected project outcome, relevant data architecture or infrastructure, approaches for validation of results, etc. (Saltz, 2015). It might be a notion that there is no need for such a methodology to be defined since; Big Data projects are often open-ended in nature. Agile methodology can be used for such projects instead. The sheer goal of finding the “value in data” is not enough. There needs to be communication between the team regarding the next steps (Saltz, 2015).

Different process methodologies have been defined in other domains. The Software Development Life Cycle (SDLC) is used in the software development domain. Optimizing business processes is used in the operations research domain, while statistical analysis is used in quantitative research. Big Data projects do not always fall specifically in these categories, although they might be similar to them. Software projects have less focus on the data aspect. A large number of extract-transform-load (ETL) processes need to be performed. Determining the relevant data sources is a crucial task in Big Data projects. This step is not a part of the SDLC. Kaisler et al. (2013) found out that “trend analysis may not require the precision that traditional database (DB) systems provide.” This shows that acceptable levels of data quality depend in most cases on data usage (Kaisler et al., 2013). Even if any software methodology was to be applied to Big Data projects, it would be difficult to determine which software methodology to use since different alternatives like waterfall or agile are in practice. Business Intelligence is another domain that deals with making effective business decisions by scrutinizing the data available. A business intelligence system that can react to unanticipated requirements also needs to be developed (Krawatzeck, Dinter, and Thi, 2015). Thus, any combined BI methodology cannot suffice a Big Data project thoroughly.

Goodwin (2011) noted that poor communication is a factor due to which 75% of corporate business intelligence projects face failure. Thus, Team Effectiveness is an essential aspect of any Big Data project. Hackman (1987) proposed a model that focuses on different factors from input to output. It is one of the most widely used models. While continuous improvement is one of the criteria in Hackman's model, a vital factor is measuring the team's performance. The model created by DeLone and McLean (1992) is based on system creation, usage, and consequences of use.

CHAPTER 2: CRITICAL SUCCESS FACTORS

For an organization every year, a large amount of information is generated regarding its employees, customers, business partners, suppliers, etc. Volume, which is one of the attributes of Big Data, is aptly named because of the vast number of data sources and the size of data generated by these sources. Big Data solutions should not only focus on the technological aspects, but also on the challenges that may occur during the project lifecycle.

CSFs are the few key areas, where "things must go right" for the business to flourish and for the manager's goals to be attained" (Bullen and Rockart, 1981, p. 7), also they are common means of assessing projects (Nixon, Harrington and Parker, 2012). Various challenges of human and organizational components of a project can be approached and tackled by understanding the related CSFs (Fortune and White, 2006).

The study of CSF for project management began in the 1960s, several lists of factors have been published where some researches have focused on specific problem domains and types of activity, and others have suggested CSFs, which can be applicable to all types of projects (Fortune and White, 2006). Some of the most studied CSFs are defined and examined in the next sections.

2.1. Critical Success Theories

A literature review on critical success factor theories led to varying conclusions by different researchers on the importance and the inclusion of factors (Anderson et al., 2006; Baladi, 2007; Delisle, 2001; Hass, 2006; Nasr, 2004; Pinto, 1986; Shao, 2006; Westlund, 2007; Wu, 2006). The theories include dynamic importance of factors theory, critical success indicators theory, integrated project planning, and control system theory, competent project manager theory, communication theory, and other theories. Professionals frequently use the principles of these theories to affect project performance. The dynamic importance of factors theory was used as the foundation for the current study. The literature review yielded conflicting evidence on conclusion validity. The discussion included validity concerns and concerns that led to excluding theories from the study found.

The concept of success factors was first introduced by Daniel who discussed it in relation to the information management crisis that was being brought about because of rapid organizational change (as cited in Fortune and White, 2006). In 1979, Rockart mentioned the concept of Critical Success Factors in the Harvard Business Review. They are termed as the crucial areas where appropriate results are necessary to achieve project success. Since these areas are of critical importance, the project manager should have the necessary knowledge to determine if progress is steady in the respective areas (Bullen and Rockart, 1981).

The study on CSF approach was established and popularized by several authors, the most relevant research work being done by Rockart (Bullen and Rockart, 1981). In his seminal paper on the topic, Rockart defined the term critical success factors as “the limited number of areas in which results, if they are satisfactory, will ensure successful competitive performance for the organization” (Bullen and Rockart, 1981). He also emphasized that these are the areas of activity in a business where constant and careful attention from management is necessary to ensure attainment of organizational goals (Bullen and Rockart, 1981). Pinto and Slevin have defined CSFs as “factors which, if addressed, significantly improve project implementation chances” (Pinto and Slevin, 1987).

Pinto (1986) began with an objective of contributing a clearer understanding of project life-cycle dynamics on critical success factors. Prior to Pinto’s research, critical success factors were primarily concepts without empirical data to support the concept within informal implementation processes. In the 1980s, the disagreement was increasing among researchers about single factor importance throughout the project life cycle.

Pinto (1986) used a survey mailed to full-time project managers and then performed data analysis related to critical success factor associations with project success at four milestones in the project life cycle. Pinto sought to establish that critical success factors are not equally important throughout the project. Pinto concluded with a critical success factor list showing a significant relationship to project success. Pinto included beta weighting to evidence the change in single factor importance during the life cycle. In the current research study, Pinto’s conclusions were used as foundational work to expand on the existing literature.

Delisle (2001) used exploratory mixed-method research involving three surveys to observe relationships among project success with Communication. Delisle focused the sample on virtual teams and establishing critical success indicators. Delisle’s research conflicted with previous established critical success factors pertaining to traditional project teams, such as project team experience, ability to troubleshoot, skills related to technology, project team culture, and tendency to take the risk. The difference in the research results compared to the established literature might be due to Delisle's use of virtual teams.

Delisle (2001) faced challenges with common terminology, the inability to establish a foundational project management theory, and the issues resulting from a sample new to the online data collection method. The respondents in Delisle's study had more experience with traditional projects than virtual projects. Delisle observed key differences between virtual teams and traditional teams. Delisle noted differences pertaining to the communication media type used and specific task leadership based on a member's expertise rather than leadership based on formal project roles.

Nasr (2004) sought to improve the existing project management systems by extending existing functionalities with an integrated project planning and control system in an effort to improve efficiencies and establish a consistent process for measuring project performance. Nasr noted standard scheduling practices with control techniques were beneficial management functions to project teams. Nasr observed limitations or deficiencies in common practices that limited the benefits.

Nasr (2004) developed a test case study as a simulation environment to measure performance with the integrated project planning and control system and to measure performance without the system. Nasr found the integrated project planning and control system benefits were more noteworthy for less experienced project managers. Nasr failed to prove the integrated project planning and control system provided any additional benefits over existing project management systems for more experienced professionals. Nasr did not prove existing project management systems are deficient or lacking functionality when compared to the integrated project planning and control system.

After conducting a project management literature review, Shao (2006) concluded three elements are key factors in determining project success: (a) a project manager with competency in project management skills, (b) a project definition that represents the project objectives, and (c) a correctly organized project. Shao examined selecting an appropriate project manager with a questionnaire founded on recommendations from the Project Management Institute regarding knowledge, performance, and personal competency. Shao (2006) used findings to build a new tool to assist with selecting effective project managers. Shao examined a single critical success factor association to project success; the research was not a comprehensive evaluation of other contributing factors, which raises internal validity concerns.

Anderson et al. (2006) concluded project communication was a success factor based on principal components analysis performed with questionnaire responses on projects. Anderson et al. focused on a single critical success factor association to project success. Anderson et al. excluded other factors, which led to internal validity concerns. The concern pertains to variables other than the predictor variable that may be responsible for the effect observed in the Anderson et al. study.

In a virtual project management study, Baladi (2007) noted contributing success factors are communication and leadership. Baladi established the conclusions with observations in the questionnaire data with a combination of t-tests, Spearman rank correlation coefficients, and chi-square statistical analysis. External validity concerns exist because Baladi (2007) limited participation to virtual project team members so bias might exist when generalizing to other project types. Baladi did not answer the current research study questions pertaining to the effect on information technology project success.

Westlund (2007) and Wu (2006) studied the project success factors associated with skilled technology resources retention and concluded a key technical resource loss before a project conclusion increases the probability of project failure. Westlund and Wu established an important factor in information technology project performance. Westlund and Wu excluded other critical success factors related to information technology project performance. The factor exclusion introduced internal validity concerns pertaining to other variables that might be responsible for the effect observed in the study.

Hass (2006) indicated information technology project success factors are a failure to integrate lessons learned failure to establish a core team, failure to create a project charter, failure to engage stakeholders, and failure to schedule a kickoff meeting. Hass presented results relevant to the current research study. The issue is Hass did not discuss the research methodology used to reach conclusions so assessment validity was not supported. Hass' conclusions were not considered as a foundation for future research.

Agirre Perez (2007) noted a risk model is a key success factor for projects with high uncertainty, such as those found in research, information technology, or aggressive product development. The Project Management Institute (2004) published a best practices guide reporting project risk management is an important factor in project success. Risk management encompasses risk identification, analysis of risk, response planning to a risk event, and monitoring and controlling risk elements (Project Management Institute). Agirre Perez's study is applicable to the current research study, but a single project management activity study raises internal validity concerns. The concern pertains to other factors being responsible for the effect observed by Agirre Perez.

Pinto (1986) established the dynamic importance of critical success factors during a project life cycle with empirical data. Delisle (2001) established critical success indicators by building research on Pinto's conclusions, but the results did not completely support Pinto's conclusions. The conflicting conclusions might be due to the focus on virtual teams by Delisle. If virtual teams affected Delisle's (2001) results, a future research study is needed to explore the effect of team types on relationships among critical success factors with project performance.

Nasr (2004) examined the potential gaps in existing project management standards but did not demonstrate existing project management systems are deficient or lacking functionality, thereby leading to limited benefits to management. Shao (2006) used a literature review to establish three factors to determine project success: (a) project manager, (b) project definition, and (c) project organization. Shao substantiated that project manager selection influenced project success.

A research by Müller and Jugdev (2012) highlighted the evolution of project success from the seminal literature to recent literature, increasing the knowledge base of project success with respect to ideas, themes, research methodologies, and the founding theories of the concept of project success. The study concludes that the concept of project success is multi-faceted. Also, some researchers in project success based their studies on organizational theories, reflecting the multi-faceted and intertwined ideas surrounding this subject. Further, Müller and Jugdev (2012) concluded that project success is a product of the combination of personnel, project, team, and organizational factors. The effective combination of these factors results in successful completion of the project. Also, project success is achieved through effective teamwork, time, cost, and scope management. Project success is also relative according to the perception measurement matrix. Different methodologies have emerged as the concept of project success evolves. These evolving methodologies involve multiple variables that work well for large, small, medium, and complex project types. Finally, the methods of measuring project success also evolve as new and robust tools, and validated and reliable instruments, are being developed (Müller and Jugdev, 2012).

A study by Sudhakar (2012) categorized the success factors for software development projects and identified the various factors in each of these categories. This categorization is a tremendous achievement in the field of critical success factors that affect software development. Before this, researchers spent more time on another aspect of success factors that influence the success of a software development project: technical, communication, and project management factors. This study, through an extensive search of the literature, used a conceptual model to identify seven success factor themes with 80 success factors sub-divided within these themes, and the first five success factors in each of these seven success factor themes are designated as the critical success factors. The selection of these 80 success factors was based on their importance in the software development discipline and their frequent appearance in critical success factor studies. Each of these success factor themes identifies five factors, which are called critical success factors. The critical success factors include communication, top management support, clear project goal, the reliability of output, project planning, teamwork, project team coordination, quality control, client acceptance, the accuracy of the output, reduced ambiguity, maximized stability, realistic expectations, and user involvement. Another major highlight of this second study is the categorization of these critical success factors into seven themes namely project management, technical, organizational, product, environmental, team, and communication.

In the study by Stankovic et al., (2013), technical factors were found to be less valid than people and process factors. Their study yielded a Cronbach's Alpha value of 0,680, 0,794, and 0,778 for the people, process, and technical factors respectively. This indicates that the people and process factors as more likely to influence software development projects than the technical factor. The model adopted by the study could report the success of all projects accurately, and it has a very high degree of operationalizing its variables. Further, it establishes that the people factors mainly involve the customer and the capabilities of the team, while the process factors involve project management and project definition processes, and these variables were evaluated based on the four success criteria of cost, time, scope, and quality.

Ahimbisibwe, Cavana, and Daellenbach (2015) systematically reviewed 148 articles and identified 37 critical success factors which were classified into three broad categories: (a) organizational, (b) team, and (c) customer factors. Under each category, the critical success factors are arranged according to their frequency of occurrence in the critical success factor literature, particularly within the traditional and agile software development methodologies. This study was carried out in four phases. The phases are a comprehensive review of the literature to identify the critical success factors for software development projects and analyzing the identified critical success factors. The other phases are differentiating the critical success factors across the different methodologies and developing a contingency fit model. This model is the first comprehensive contingency fit model in the study of critical success factors, and it distinguishes clearly between both the traditional and the agile methodologies. The contingency fit model employed by this study helped in developing a model that can determine the degree of that influence the critical success factors has on project success.

The lack of proper solution designs or architecture for Big Data problems is among the prime technical problems. The technology to be used needs to be customized according to the type of analysis that is to be done via the project. Storage of data should be taken care of from the initial stages of the project. Data might be needed to be transformed into another form, to make it more structured, and make it a better fit for the business requirements. There is also a possibility of information loss during the process of transforming the unstructured data into a more structured format (Gopalakrishnan et al., 2012; Cuzzocrea, Song, and Davis, 2011). While merging the data from different sources, other concerns like security and privacy of data need to be taken into consideration as well. Access control mechanisms should be implemented to allow access to specific data to specific people depending on their role. Data spillage is an important concern especially when cloud-based platforms come into the picture. Storage, retrieval, and processing data on such cloud-based systems have a huge overhead especially when security comes into the picture (Gao, Koronios and Selle, 2015).

The research of Saltz and Shamshurin focuses more on the need of people, process, and technology context of Big Data projects. The authors put forth six categories for the CSF for Big Data projects. They are listed as follows (Saltz and Shamshurin, 2016): Data (access, security, ownership), governance (culture, management, performance), process (project management and change management), objectives (well-defined goals), teams (structure and skillset), tools (technical aspects).

The importance of change management, as well as the inclusion of procedures and policies for data, is also stressed by Wamba et.al, (2015). Different authors have proposed various challenges in adopting Big Data technologies. These challenges if coordinated with critical success factors can be done by Saltz and Shamshurin (2015), three categories were identified by Yeoh and Popovic (2016) namely, Organization, Technology, and Process. On the same lines, Evers identified the categories as Organizational, Performance, and Technical (Saltz and Shamshurin, 2015).

Chen et al. identified seven critical success factors without categorizing them separately (Chen et al., 2016). These are customer-centric focus, pre-project value discovery, strong business need, talent planning technology infrastructure, top management involvement and vendor contract management.

The literature does not provide any empirical studies or publications that have tied CSFs of Big Data projects and illustrated their correlation. There is, however, one conceptual model that looks at Big Data projects. Halaweh and Massry (2015) presented a conceptual model for research. The 5 dimensions from Wamba et al., (2015) are data policies, technology, and techniques, organizational change, and management, access to data and industry structure. squarely focus on utilizing the field of Big Data as we explore the relationship with successful Big Data projects.

Figure 6: Conceptual Model of Halaweh and Massry

Source: Halaweh and Massry (2015)

The model above focuses on the 3 Vs for and utilizes challenges, failure and success criteria and obstacles needed for successful implementation of Big Data that exists in the literature. The model is completely based on the review of the literature, is generic and conceptual in nature. The two future directions listed by both authors (a) use quantitative research methods to test model and verify the validity of the assumptions and (b) evaluate model by applying qualitative methods to semi-structured interview experts who work in different sectors to develop or extend the model that affect Big Data implementation. This research study uses both methods (a) and (b) to develop and test statistical relationships of the model.

Organizing the literature on critical success factors involved categorizing findings on causes for poor project performance related factors. Next sections include the literature on several critical factors in the light of CSFs of Big Data projects.

2.2. Human Capability

The critical success factor of human capability, according to the literature, refers to appropriate and necessary human resources that will effectively run and manage the project and its available resources (Browning and Ramasesh, 2015; E Silva and Seixas Costa, 2013; Kuen and Zailani, 2012).

Boehm and Turner (2005) stated that the choice of technology and other technical resources is the prerogative of the project manager and subsequently the project personnel. Despite the broad approaches available, the project personnel tend to remain within their comfort zones by choosing methodologies they are conversant with rather than choosing the appropriate methodology (Boehm and Turner, 2005). This choice is one significant disadvantage of using unqualified personnel (Howell, Windahl, and Seidel, 2010).

2.3. Organizational Capability

Latonio (2007) executed a phenomenological study design by interviewing 20 project managers from 16 industries. Latonio established project success factors related to leadership influence, management effect, project success criteria consideration, supporting values, communication, and organization leaders’ commitment to the project. Ikeda (as cited in Jedd, 2007) reported project failure is a combination of implementing a project structure without consideration to the corporate culture and failing to tie organizational strategy with the project priority.

Dinsmore and Cabanis-Brewin (2006) identified the inadequate project managers as the primary cause of the project. Dinsmore and Cabanis-Brewin denoted the condition stemmed from an inadequate organizational incentive to transition an individual from a technical expert to a project manager. Dinsmore and Cabanis-Brewin also stated the insufficient definition of the project manager role leads to a single individual assigned with too much responsibility.

Wysocki (2004) concluded organizations implementing a project methodology might not benefit from decreased project failure rates if the organizations do not take measures to protect the investment. Wysocki found project teams are encouraged to use the accepted methodology prevalent in an organization, but often the individuals do not fully embrace the change introduced by the methodology or use the methodology tools as intended. Organizational leaders might consider a project methodology implementation an investment and might take measures to protect the investment by ensuring its proper use after implementation to realize higher success rates with projects (Wysocki).

Sidenko (2006) concluded maintaining technology project success was related to project management maturity models, standard project practices, and project management tools being used by personnel in organizations. Ojiako, Johansen, and Greenwood (2008) used a grounded theory study to examine two case studies from major industries and the effect on project success when the project and business objectives did not align. Ojiako et al. concluded a universal success factors checklist is not possible because variables differ due to size, distinctiveness, industry, perceived complexity versus real complexity, and stakeholder composition.

Henry (2004) observed organizational processes in the technology department and the business departments affected project success due to knowledge transferability and project governance alignment. Woodward (2007) advocated project failures are due to (a) inadequate sponsorship at the executive level, (b) unrealistic deadlines from management, (c) incompetent project managers applying wrong project management methodology, (d) insufficient end-user involvement, (e) poor requirement documentation, (f) inadequate change management, (g) insufficient communication, or (h) inadequate cost and schedule estimations.

Connelly and Canestraro (2007) studied information technology projects between governmental agencies. Connelly and Canestraro observed project failure was related to fundamental organizational issues and behavioral issues. Critical success factors included elements pertaining to organizational factors and behaviors factors. Zhao (2007) concluded both organizational behaviors and project management are contributing factors for technology upgrade projects.

Zhao used a mixed-method approach to collect data from 15 project manager interviews and observed the following contributing project success factors: (a) establishing a vision for the organization, (b) establishing a communication plan, (c) minimizing customization, (d) obtaining support from external sources, (e) establishing project management techniques, (f) establishing executive-level support, (g) setting up training, and (h) securing end-user involvement. Zhao collected survey data for use in statistical analysis to establish relationships among contributing factors with project success. Zhao concluded contributing factors indicated a varying level of importance at different points in the project life cycle.

In accordance with authors of project performance textbooks, similar results were observed in a public opinion poll administered by the 2007 Computing Technology Industry Association that received more than 1,000 responses (Deliverables, 2007). The survey respondents chose different reasons for the leading cause for project failures: (a) 28% identified poor communication, (b) 18% selected insufficient resource planning, (c) 13.2% selected unrealistic schedules, (d) over 9% choose poor project requirements, and (e) over 6% selected insufficient stakeholder buy-in. The remaining responses were undefined project success or closure criteria, unrealistic budgets, insufficient or no risk planning, and an inadequate control process or change process.

2.4. Technical Capability

This critical success factor concerns the availability of adequate technology, adequately equipped personnel, and the provision of other technical resources needed to complete the project (Ahimbisibwe et al., 2015; Kuen and Zailani, 2012; Pope-Ruark, 2015). Sauser, Reilly, and Shenhar (2009) and Murad and Cavana (2012) noted that many IT projects fail due to inadequate technology, ill-equipped personnel, and lack of other technical resources.

2.5. Project Management

Project managers must focus on eliminating distractions from team members (Indelicato, 2007). For example, on a project with constrained resources, a project manager cannot allow resources to become sidetracked by outside distractions (Indelicato, 2007). Parchoma (2007) studied projects to integrate technology into instruction. Parchoma discovered project performance was affected by adequate time to balance project commitments with other responsibilities. Project managers are responsible for setting limits in addition to disengagement through communication to maintain team members' focus on the team members' area of expertise to achieve the highest project resource efficiencies and effectiveness (Fretty, 2007).

Leach (2005) noted project success is achieved by strict product scope management through a change request process. Each request is processed up front and assigned an impact estimate to the budget and an impact on the schedule (Leach). The request is also considered against any additional project risk introduced by the change (Leach). According to Kerzner (2003), technical failure and insufficient risk management by the project manager cause project failures.

Lewis (2007) consistently found project failure causes were related to inadequate project task planning by the project manager. The planning deficiency led to inevitable rework and wasted time on trivial distractions (Lewis). Lee and Hirshfield (2006) found failures in health-care software implementations stemmed from poor up-front planning fueled by anxious team members, a push to realize a return on investment, and conflicting priorities. Sauser (2005) concluded project success relates to effective leadership skills, efficient management skills, and a project manager's technical competence. Sauser noted the project manager must accept responsibility for the project vision, execution, and resulting product. Sauser based conclusions on the observations from case study research.

Kendrick (2003) identified three reasons for project failure: (a) infeasible technical functionality, (b) unrealistic schedule expectations, and (c) inadequate planning by the project manager. Frame (2003) identified project failure as resulting from three primary sources: (a) organizational elements, (b) inadequate requirements, and (c) inadequate planning or control by the project manager. The report published by the U.S. General Accounting Office (1997) after reviewing the U.S. Department of Energy included four major causes of project failure: (a) unclear product scope or considerable product scope change, (b) incremental project funding, (c) misaligned incentives, and (d) insufficient contractor overseeing.

Peslak, Subramanian, and Clayton (2008) evaluated commercial off-the-shelf implementation projects. Peslak et al. concluded through confirmatory factor analysis the key product use factors are preparation, training efforts, and efforts pertaining to performance and usefulness. Peslak and Stanton (2007) observed 18 teams using exploratory factor analysis to determine project success factors include emotions and establishing processes by involving the team and related personnel.

Liemi (2004) and Fan (2007) both concluded similarly in separate studies that knowledge management is a key success factor to project management. Liemi used a project management survey to observe a positive relationship between project management and knowledge management techniques. Fan established knowledge management technologies, such as data warehousing and data mining, improved the project performance of construction equipment management.

2.6. Project Definition

Project Definition enumerates the goals, purpose, and the focus of the project. The properly defined project helps the project team to focus, be on target, be extremely committed, and operate in one accord to achieve project objectives (Kuen and Zailani, 2012). The Project Definition is one single critical success factor that the literature has described as very significant for success in every phase of the project lifecycle (Müller and Jugdev, 2012).

Project schedule/plan is also a part of the Project Definition and it should be planned clearly before the project starts. Project schedule/plan is a comprehensive outlook of the necessary processes, procedures, and all resource requirements including financial and human resources that will aid in the successful completion of the project (Dezdar and Ainin, 2011; Iamratanakul et al., 2014; Kuen and Zailani, 2012; Moohebat et al., 2011). According to Dezdar and Ainin (2011), the project schedule/plan must detail the project activities including the appropriate timelines, appropriate human resources, monetary resources, and all other necessary resources that will lead to the successful completion of the project.

2.7. Change Management

Big Data projects require a change process (Hallikainen, et al., 2006). It can be viewed from two perspectives: implementing change and adapting change (Garg and Singh, 2006). When a Big Data implementation project is initiated by top management, top management must ensure that their staff will and can adapt it. An environment which change can be implemented is required for the Big Data implementation (Calvert and Carroll, 2005). If their staff do not aware the change or do not adopt the change, the result of the implementation may not be as expected. In order to make the change effective, it requires change management. Al-Mashari and Zairi (2000) stated that change management facilitated the insertion of newly implemented systems, processes and structure into the working practice, and dealt with resistance. Kemp and Low (2008) indicated that change management was required to prepare users for the introduction of a new system, to reduce resistance towards the system and to influence user attitudes towards the system. These objectives ensure the acceptance and readiness of the new Big Data project output, allowing the organization to get the benefits of its use (Esteves and Pastor, 1999).

Kemp and Low (2008) proposed a range of change management activities, such as communication, project championship program, training, users' involvement, and phased implementation. All of them were regarded as critical success factors in other studies. In order not to duplicate, change management was not considered as a critical success factor. Nevertheless, the construct of change management is still the backbone of the current study because it is believed that Big Data projects require a change process. Sarkis and Sundarraj (2003) indicated three issues should be addressed for change management, namely user expectation, user involvement, and user satisfaction

2.8. Communication

Communication is the art of providing an appropriate medium for seamless interaction and collaboration among all stakeholders (Fesenko and Minaev, 2014; Kisielnicki, 2011; Sidawi, 2012). Kuen and Zailani (2012) stated that communication is a critical factor that affects the successful completion of an IT project. Effective communication increases knowledge, identifies risks, eliminates or minimizes unproductive activities, reduces errors, and helps to create ideas that could lead to the successful completion of the project

(Kuen and Zailani, 2012). Communication within the project team is one of the most critical success factors. Browning and Ramasesh (2015) reiterated that collaboration between the project team, the different groups, all stakeholders, and all operations is essential to the successful completion of the project. Additionally, effective and appropriate communication between all parts of the system including personnel plays a significant role in the success of a project.

Besides communicating within the project team, communication with other members in the organization is crucial. Effective communication between the project team and the end users, particularly during the analysis and design phases, is crucial to producing a project successfully. Further, effective communication could be achieved through a seamless collaboration between the project team and the end users throughout the developmental stages of the Big Data project (Mavetera and Kroeze, 2009). Finally, the effective use of communication tools and techniques is critical to the success of Big Data projects. The tools and techniques include the use of media to inform the project team and all stakeholders of the project's progress, brainstorming meetings, and the use of the pair programming technique (Mavetera and Kroeze, 2009).

It means that information is not only shared between project team but also communicated to the whole organization the results and the goals in the implementation phase. The communication effort should be done on regular basis (Esteves and Pastor, 2000; Sternad and Bobek, 2006). Top-down, bottom-up and horizontal communications are required in the course of Big Data implementation. It is important to understand the differences in perceptions of project team members and non-project team members in designing communication mechanisms (Amoako-Gyampah, 2004). Effective communication is a key element which helps disseminate new information, challenges or opportunities to all parties involved (Muthusamy et al., 2005). Expectations and goals must be communicated among stakeholders in all levels of the organization. Stakeholders must understand the capabilities and limitations of the Big Data project. Otherwise, the Big Data project may fail to meet stakeholders’ expectations (Nah and Delgado, 2006). Zhang et al. (2005) further indicated that an open system culture should be encouraged. People within a closed system would think they were going to be constrained by the Big Data project, which inevitably led to resistance to the Big Data project.

Communication breakdown is one of the uncertainties in Big Data projects. It is sometimes unavoidable due to languages or technical jargon used. To avoid it, Loh and Koh (2004) suggested that clear instructions and messages should be given all the time. Educational workshop and training can enhance users' knowledge and eliminate or minimize unnecessary communication breakdown.

In short, change management requires user involvement and participation. User involvement and participation require teamwork and effective communication. So teamwork and communication is the first critical success factor identified for the current study.

2.9. End-User Acceptance

End-user acceptance, which is the extent to which the client accepts and uses the developed project output (Kuen and Zailani, 2012), is a very vital success factor in IT projects. End-user acceptance is the acceptability and usability of the product by the clients, and this determines if the project is a success or a failure (Kuen and Zailani, 2012; Müller and Jugdev, 2012; Ofori, 2013; Sudhakar, 2012). Also, the literature reveals that frequent communication and consultation with the end user to get feedback, particularly about meeting the needs of the customers, is essential to a successful implementation of the Big Data project. Further, it is essential that the client should be conversant and be in agreement with the project success criteria from the initial stages of the project (Kuen and Zailani, 2012; Pope-Ruark, 2015).

2.10. Training

Training is one of the critical success factors in IT projects (Bagchi et al., 2003; Yang and Seddon, 2004). Lack of training and education are the number one IT implementation problem in small and large manufacturing firms (Duplaga and Astani, 2003). In a study about players and activities across the project life cycle, Somers and Nelson (2004) concluded that user training was important throughout the implementation cycle. Training is regarded as important events which must be arranged in consideration of the implementation phases. Calvert and Carroll (2005) pointed out that the timing and scope of training were logically related to the phases of the implementation project. In the planning phase, project team members should go off to train on the project output (Clinton and Lummus, 2000; Mäkipää, 2003). Inadequate training will cause a large number of errors and problems in testing the Big Data project. Cutting the time allotted to testing and training increase the chance of failure (Markus and Tanis, 2000). End-users training is typically the last activities in the project phase (Markus et al., 2000). Besides formal training, other mechanisms such as help desk, online help, knowledge management systems, communities of practice and establishment of power users, must be established (Calvert and Carroll, 2005).

As discussed, training is a necessary event in Big Data implementation, but whether it can lead to a success is not determined. Antonacopoulou (2001) indicated that training could not be assumed to produce learning. Training is based on control and conditioning of individuals' understanding, whereas learning is about broadening and liberating understanding. In training, the trainer can train users how to use the Big Data tools with demonstration data of some scenarios. Users are required to learn how to apply the skill and knowledge to other scenarios which are not covered in the training. If users are not able to apply the skill or knowledge to other scenarios, it can say that users complete the training but the training is not effective because they do not learn what they need to learn. Besides technical and operative knowledge, training should also cause behavior change (Laoledchai, Land and Low, 2008).

Calvert and Carroll (2005) used the term "change management" to replace "training strategy" because it takes a holistic view of training in a Big Data environment. Change management should ensure users learning what they need to learn. Once users are able to handle the tool themselves, resistance resulted from fear of disruption will be reduced. In the learning processes, users will be familiar with the new tools, new processes, new relationships and structures, resistance resulted from long standing organizational traditions and work processes will be reduced. The effectiveness of training is one of the primary concerns in Big data projects, not training itself.

2.11. Top Management Support

Top management support, a critical success factor in all types of investigated projects, essentially relates to the unflinching support of senior management to the success of the project by providing every needed support necessary to complete the project (Elbanna, 2013; Garrett and Neubaum, 2013; Lee, Shiue, and Chen, 2016). Top management support includes but is not limited to the provision of adequate financial assistance and all other necessary resources for the successful completion of the project (Elbanna, 2013). Nah et al. (2003) indicated that top management support influenced both commitment to change management and commitment to resources, which were necessary factors for success in Big Data project. The implementation project should be identified as a top priority which encourages the entire organization to focus on the project and motivates the project team and users to learn the Big Data tool and truly participate in the project (Wang, Klein and Jiang, 2006). Top management must help project team members move into a high-performance team and then assist them to move from teamwork to team learning. Teamwork can create synergies and get the problem solved. However, team learning encourages the team members to learn from others, help others learn and learn about working with each other (Nagendra, 2000).

Top management should also allocate the necessary resources to the Big Data project (Nah et al., 2003). Jafari et al. (2006) found out that allocating necessary resource was the most important duty of top management in a Big Data project. The attitude of the top management to the Big Data project determines the number of resources allocated (Nah et al., 2003).

Dedication from the executive level is significant during all activities associated with Big Data implementation and upgrade (Nah et al., 2001; Wenrich and Ahmad, 2009). Without top management support, there is little hope for it. This is especially important in the early stages of an implementation project (Akkermans and Helden, 2002). Top management must define objectives of the Big Data project in order to give the project team and users a clear business plan and vision to steer the direction of the project (Loh and Koh, 2004; Francoise et al., 2009). Also, they should paint a picture of where the organization will end up and portraying the anticipated outcomes after the Big Data project (Martin and Huq, 2007).

Although top management support is widely regarded as an important factor in the literature of Big Data, Nah et al., (2007) indicated that it acted more like an "enabling" rather than a necessary factor for the projects in developing countries. In their study, top management support did not impact the success of projects in developing countries. From the process point of view, top management commitment is a necessary factor that top management must make the decision to acquire and to implement the Big Data project.

Without their approval, the project phase will never happen. Nah et al. (2007) also indicated that top management might be necessary for the completion of a Big Data project but might not directly affect the effectiveness of the system. Kamhawi (2007) also found that top management support was not significantly related to both project and business success in a regression analysis but it had a significant relationship with the Big Data success in the correlation analysis. The contrast in results means that although top management support is related to Big Data success, its interaction behavior with the other critical success factors is not significant in relation to success dimension (Kamhawi, 2007).

Based on the literature reviewed, top management is a necessary and important factor for Big Data projects. However, why does top management have these supportive behaviors? The linking of the Big Data project with enterprise strategy is one of the elements strongly influencing the top management behaviors. Big Data projects are perceived by top management as a means by which an organization can complete its strategic goals which can be included both tangible and intangible objectives. Intangible strategic goals can be organization development and growth, customer satisfaction or information availability. Tangible strategic goals may include operating cost reduction or an increase in profitability (Soja, 2008). A Big data tool is a strategic tool to introduce changes to organizations for particular strategic goals, such as standardization, competing against competitors, winning market shares and sustaining competitive advantages (Kraemmer et al., 2003; Jafari et al., 2006; Baray, Hameed and Badii, 2006; Olugbode et al., 2008; Baray, Hameed and Badii, 2008). From a strategic point of view, the success of a Big Data project can refer to the increased value of the business from usage of the Big Data tools (Nah et al., 2007)

2.12. Troubleshooting

Troubleshooting, as a critical success factor, is used as a general term for troubleshooting, monitoring and feedback, and end-user consultation activities. Troubleshooting mainly concerns the capability to promptly manage uncertainties and inherent issues developing during the life cycle of the project (Ahmad et al., 2012; Kuen and Zailani, 2012). Due to unforeseen circumstances, situations may develop, hence the software development team should be ready to tackle emerging crises and arising deviation from the initial plans

(Kuen and Zailani, 2012). Also, the project team should be versatile in the concept of risk management to troubleshoot effectively should unanticipated incidences arise (Kuen and Zailani, 2012).

The monitoring and feedback is another critical success factor for Big Data projects. This construct allows for prompt and timely intervention in the event of any adverse contingencies that may affect the success of the project (Kuen and Zailani, 2012; Shatat, 2015).

Client consultation primarily details active consultation with all stakeholders and incorporates all necessary functions that will aid the usability of the software product (Ofori, 2013; Sudhakar, 2013). Also, client consultation affords every stakeholder the opportunity to provide input particularly during the initial stages of the project management and to be informed of the progress of the project (Kuen and Zailani, 2012). In addition, stakeholders tend to embrace the project output since they have been involved during its development (Ahmad et al., 2012).

2.13. Miscellaneous

In addition to the potential influences of Big Data projects, other variables have been shown to relate to the success of projects. Project size is one of these variables. The Standish Group (2010) reported that projects were completed on time, within budget, and with the required functionality only 4% of the time for new application development, 30% of the time for package applications, and 53% of the time for application modernization projects (i.e., software updates). Ajila and Wu (2007) found that project success, defined as completing the project within the planned timeframe, was higher for smaller organizations. For the third component of the iron triangle, quality, small organizations again performed better, with projects averaging 74.2% of their originally-intended features, compared to only 42.0% for projects developed by large organizations (Standish Group, 1995). Work experience has also been showing to predict project success (McHaney, White and Heilman, 2002). For example, Müller and Turner (2007) showed that older project managers with more years of managerial experience were more likely to lead projects that concluded successfully than younger project managers with fewer years of managerial experience. In summary, both years of management experience and project size have been shown to be key variables in predicting project success. Years of experience in IT have also been shown to predict accuracy in costing and scheduling tasks (Henry et al., 2007).

2.14. Project Success

Project success is a multidimensional variable which can be broadly categorized in terms of efficiency and effectiveness of the outcome of the project (Ika, 2009; Ika, Diallo, and Thuillier, 2012). Efficiency describes the success of the project in terms of the triple constraints of the Project Management Institute (PMI), namely time, cost, and scope. Effectiveness describes the outcome in terms of achieving project objectives, business objectives, and social and environmental goals (Howsawi, Eager, Bagia, and Niebecker, 2014; Ika et al., 2012; Müller and Jugdev, 2012; Palcic and Buchmeister, 2012; Rolstadas et al., 2014). While there is no universally agreed-upon definition of project success, most do agree that success “is in the eyes of the beholder” (Müller and Jugdev, 2012).

The literature review on project success revealed different conclusions to indicate researchers and public opinion disagree on leading project success predictors. The current research study’s objective was to provide insight into the critical success factors of Big Data projects and Big Data project success. The secondary purpose was to determine the relationship between CSFs and test and validate the scale and the research model statistically. The SEM established a validation method of Big Data project success by recognizing the degree of relationship between 5 contributing success factors and Big Data project success.

2.15. Current Gap

There are many unanswered questions about big data. These questions range from attempting to define it, to asking how it will help decision makers make better decisions, how to effectively govern the immense volumes of data, and how to protect customer’s privacy. There is agreement that the sheer volumes of big data will require cutting-edge technology to maintain it and new analytical skills for it to be effectively used.

The various implementations, theories and proof of successes has been singular (tied to single instance within the organization), has only dealt with showcasing an area of focus (trying to solve one problem and not organizationally prevalent) and has not been replicable or able to pass on success/learnings to other aspects of the organization.

The need for successful project management is increasing as projects are used more to achieve operational goals in various organization types (Hyvari, 2006). Project success might link to national security in some government agencies, such as Jones' (2007) project to implement a passenger tracking system. Leaders of modern organizations using projects to achieve operational goals have a potential opportunity to reduce costs using recommendations from the current research study. Lewis (2007) noted approximately 30% of development cost is linked to rework of previously completed tasks.

Woodward (2007) and Ildefonso (2007) both concluded that personnel who follow best practices experience dramatic increases in information technology project success. This study contributes to the literature by summarizing and categorizing factors recognized in best practices.

Hyvari (2006) conducted a literature review and discovered disagreement in the project management literature on what constitutes a successful project within an organization. Shenhar and Wideman (2000) found the same disagreement pertained to defining project success in the business literature. The literature includes material covering a wide concern for project success within organizations but also wide disagreement on how to measure project success.

Since the 1960s, researchers have contributed efforts to define a single comprehensive factor set to predict project success, but have consistently disagreed on one or more factors (Cooke-Davies, 2002). Pinto and Prescott (1988) indicated research prior to 1988 was theory based without empirical data. Disagreements in the literature on a single factor set might result from a lack of empirical data. The current research study provides empirical data to the body of knowledge pertaining to associations between 5 critical success factors and Big Data project success.

There are a lot of challenges that still exist with Big Data. There is the issue with dealing with heterogeneity, inconsistency and incompleteness, varying scale, timelessness, privacy and data ownership and visualization and collaboration. (Jagadish et al., 2014)

In recent times, the research on Big Data has been always concentrated toward creating better algorithms and designing robust data models (Saltz and Shamshurin, 2016). However, not much work has been done regarding finding out the best methodology for executing such projects (Ahangama and Poo, 2015; Saltz, 2015). The exploratory nature of Big Data projects demands a more specific methodology that can handle the uncertain business requirements of such projects (Saltz, 2015). According to a survey carried out by Kelly and Kaskade (2013), "300 companies reported that 55% of Big Data projects don't get completed and others fall short of their objectives." The reasons for such project failure can be identified at the beginning of the project or can be reduced at a later stage by some coordination methodology (Saltz, 2015).

A well-defined Big Data analysis project methodology would help to address different issues like roles and responsibilities of team members, project stakeholders, expected project outcome, relevant data architecture or infrastructure, approaches for validation of results, etc. It might be a notion that there is no need for such a methodology to be defined since Big Data projects are often open-ended in nature. Agile methodology can be used for such projects instead. The sheer goal of finding the "value in data" is not enough. There needs to be communication between the team regarding the next steps (Saltz, 2015). Different process methodologies have been defined in other domains. The Software Development Life Cycle (SDLC) is used in the software development domain. Optimizing business processes is used in the operations research domain, while statistical analysis is used in quantitative research. Big Data projects do not always fall specifically in these categories, although they might be similar to them. Software projects have less focus on the data aspect. A large number of ETL processes need to be performed. Determining the relevant data sources is a crucial task in Big Data projects. This step is not a part of the SDLC. Kaisler et al. (2013) found out that "trend analysis may not require the precision that traditional DB systems provide." This shows that acceptable levels of data quality depend in most cases on data usage (Kaisler, et al., 2013). Even if any software methodology was to be applied to Big Data projects, it would be difficult to determine which software methodology to use since different alternatives like waterfall or agile are in practice. Business Intelligence is another domain that deals with making effective business decisions by scrutinizing the data available. A business intelligence system that can react to unanticipated requirements also needs to be developed (Krawatzeck, Dinter, and Thi, 2015). Thus, any combined BI methodology cannot suffice a Big Data project thoroughly.

Currently, Big Data research is concentrated on enhancing data models and algorithms; however, the best approach to execute projects must also be studied. Complicating the situation is that Big Data projects are exploratory in most cases, and accordingly, the projects lack clear business requirements with subsequent results not easily validated (Saltz and Shamshurin, 2016). Moreover, teams performing data analysis and data science work operate in an ad hoc fashion where a trial and error process is used to identify the right tools and accordingly involves a low level of process maturity (Saltz and Shamshurin, 2016).

In the 1970s, the concept of Critical Success Factors was introduced and can be defined as elements that are essential to execute the project successfully. Many studies – as discussed in the literature review – throw light on various critical success factors identified and validated for Big Data projects. Ojiako et al. (2008) concluded a universal set of success factors is not possible because contributing factors differ due to size, distinctiveness, industry, perceived complexity versus real complexity, and stakeholder composition. These critical success factors have been categorized so far into generic groups such as People, Process, Technology, etc. A categorization of these critical success factors in a statistically tested model is a gap that needs to be bridged. The practical usage of this study can help organizations to identify factors contributing to the success or failure of Big Data projects and also examine the relationships between CSFs.

The gap, characteristics and the limited amount of existing data, framework and variables exist concerning successful Big Data projects we reviewed further lends to our research questions, *"What are the critical success factors Big Data projects?"* and *"What are the relationships among the critical success factors?"*. Furthermore, the researcher reasons that organizational effects can lead the way for successful Big Data projects. Next, to investigate these research questions, the researcher will utilize three-part mixed methods study utilizing constructivist grounded theory by conducting semi-structured interviews and Delphi study with industry experts on successful Big Data projects. Then perform qualitative analyses to identify variables to measure via CATI survey and finally perform quantitative analyses to answer our research questions and proving/disproving our hypotheses.

It is also very clear from literature and practitioners of the field that Big Data is here to play a role in our future (Gamage 2014; Burg 2014; Allouche 2014; Halaweh and Massry 2015; Wamba et al., 2015; Wixom et al., 2014; Xu et al 2015; Chen et al 2012; Forrester 2012). For that purpose alone, it is imperative that we learn, take advantage and realize its potential to transform entire business processes (Wamba et al., 2015).

Hence, this study focuses on identifying the key areas – also called “Critical Success Factors” – essential for achieving project success in Big Data projects. A review of the literature indicated a gap exists in the project management literature and the business literature pertaining to a comprehensive factor list to support predicting project performance (Cooke-Davies, 2002; Hyväri, 2006). Associations between each of 5 critical success factors to Big Data project success have not been consistently established with empirical data in the literature. The current research study contains new knowledge regarding establishing relationships between the 5 critical success factors and Big Data project success.

The main purpose of this research is to build on the current diverse literature around Big Data by contributing discussion and data that allow common agreement on definition, characteristics, and factors that influence successful Big Data projects. The research questions being investigated are based on the argument establishing Big Data be used as a tool for the organization by which to develop and create efficiencies enterprise-wide. The researcher explores the following question with this research, *“What are the critical success factors that impact perceived project success in Big Data projects?”* and “*What are the relationships among the critical success factors?”*. As part of these research questions, CSFs and several hypotheses mentioning relations between the CSF will be examined. The relations between CSF will be visualized and tested in a relational model via Structural Equation Modeling (SEM).

To accomplish this, the research will follow the below five-step process:

1. A comprehensive review of the literature is conducted.

   The literature review includes relevant research regarding such critical success factors that are validated in previous studies. Several different case studies and theoretical discussions enlist success factors regarding Big Data projects. The study compiled these critical success factors as provided in the literature regarding Big Data projects. Notable success factors for Big Data projects were compiled from literature such as case studies, theoretical observations or experiments.

2. The research identifies the current gaps, definitions and existing variables from the literature regarding Big Data projects and CSFs.

3. The research employs a three-part mixed methods study based on grounded theory.

The research parts are: (a) Qualitative: Semi-structured interviews and Delphi study with experts on successful Big Data projects, performing qualitative analyses to identify factors and answer our qualitative research question, inducing a model to measure via a CATI survey and (b) Quantitative: performing quantitative analyses to test the model, answer our quantitative research question and proving/disproving our hypotheses.

4. Next, the research compiles a list of CSFs that impact success in Big Data projects from our quantitative tests.
5. Finally, the research outlines the findings.

Different challenges are encountered at an organizational level when implementing Big Data projects (Saltz, 2015). To deploy and exploit Big Data in an optimal manner, it is necessary for the organization to pay more efforts in managing these projects more efficiently. The literature review uncovered several research efforts on project success, performance and studying relationships to one or two critical success factors, but the existing literature did not contain a statistically tested research model, empirical data and statistical findings to answer the research questions of the current research study.

CHAPTER 3: RESEARCH METHOD

From the previous chapter, we concluded the following: (a) little existence of quantitative empirical study on CSF of Big Data projects in the literature, (b) very little research on theories, makeup, frameworks or definition around variables investigating CSFs of Big Data projects and (c) due to its infancy, emergence of new challenges that have been brought to light primarily due to lack of standardizations and accountability (Halaweh and Massry 2015). Challenges such as privacy and security have been clearly documented in the literature (Wamba et al., 2015; Halaweh and Massry 2015).

Given the above and the early nature of this concept, the main purpose of this research is to investigate CSF in Big Data projects driven primarily around premise of the Big Data experts' arguments regarding Big Data projects, as a field of study, leading such innovations and being strategic (Galbraith 2014, Church and Dutta 2013, McAfee and Brynjolfsson 2013, Wamba et al., 2015, Halaweh and Massry 2015). There have been calls suggesting that CSF is strategic (Jelinek and Litterer 1988; Head 2009) and experience with many methods that tie in with Big Data historically as evidenced by Weisbord (2012) analysis of CSF history.

Unfortunately, a very limited amount of existing data, framework and variables exist concerning successful Big Data projects. It was, therefore, important to formulate a method that would allow us to collect data, review, analyze, deduce a model, formulate a theory and finally test the phenomenon statistically.

The idea of rigor was especially central to this study of Big Data project success because (1) the consequences of project success are less observable in the enterprise (since other parameters may yield to positive or negative performance at the end of the project), and harder to identify lack of which factors could ruin a Big Data project, and (2) there are fewer phenomena to study on the whole, thereby, ensuring that the proposed study would illuminate the core philosophical tenets of the process of Big Data project success. The focus of the study was on the process, context, and implementation of Big Data projects, as guided by a social constructivism lens, and the grounded theory approach (Charmaz, 2006) focusing on the theoretical orientation of Big Data professionals' views and perspectives.

3.1. Research Problem

There has been limited empirical research on organizational factors that relate to Big Data (LaValle et al., 2011; Bean and Kiron, 2013). Even though there has been some empirical work on the technical, organizational, and individual factors related to Big Data adoption and success (Uğur and Turan, 2018; Al-Qirim et al., 2017), a gap exists in terms of understanding the critical success factors (CSFs), such as organizational size and top management support, that relate to Big Data project success. Previous studies have focused primarily on the technical and individual factors that relate to Big Data adoption. Sim (2014) acknowledged this gap and suggested that organizations should be aware of the important factors for Big Data success.

Critical success factors have not been investigated as a group of organizational factors that relates to Big Data success. However, researchers have examined critical success factors as an important factor during IS implementations (Davis, 2014; Dong, 2008; Tarhini et al., 2015). Several authors have conducted quantitative studies of how critical success factors support relates to specific technologies, including service-oriented architecture (SOA) (Maclennan and Van Belle, 2014), accounting information systems (Anggadini, 2015), healthcare information systems (Hung et al., 2014), and ERP systems (Dong, Neufeld, and Higgins, 2009; Palanisamy et al., 2010; Tarhini et al., 2015).

The lack of critical success factor sources can doom an IS project to certain failure. Elbanna (2013) argued that critical success factors have to be consistent and constant during a project implementation, otherwise the project could fail. Although IS success was studied as it relates to IS implementations, critical success factors have not been discussed as it relates to Big Data projects. Some critical success factors are significant for both IS projects and also for Big Data projects. Top management support is one of these common critical success factors (Barclay, 2015; 2016; Young and Poon, 2013). Young and Poon (2013) suggested that top management support is nearly always necessary for an IS project to be successful because the top management team can influence the success or failure of a project. Conversely, Young and Jordan (2008) argued that project planning, user involvement, and project methodology are not critical success factors for an IS project. But these factors may be critical for a Big Data project. Big Data implementations vary from traditional IS projects in terms of requirements as; multi-disciplinary teams, agile development with frequent business user check-points, data profiling, visualization, non-deterministic outcomes, change management, optimizing resource management.

There has been little research conducted related to IT professionals and big data. Specifically, to our knowledge, there have been no studies to determine critical success factors of Big Data projects and examine the relationship between the factors. This research can help organizations in general to identify factors that impact success – as perceived by practitioners and professionals – on Big Data projects.

3.2. Research Design

In their paper, Amberg, Fischl, and Wiener (2005) have listed the following to be the most frequently used methods to identify relevant CSFs: action research, case studies, Delphi technique, group interviews, and literature review, multivariate analysis, scenario analysis, structured interviewing.

This research is exploratory in nature. The best approach for such a study was mixed methods utilizing Constructivist Grounded Theory. Mixed methods allow for the integration of qualitative and quantitative data within a study to provide a more complete analysis of the research problem being investigated (Creswell and Plano Clark, 2011). It allows for, especially for an early concept, data to be built and further explored using a secondary method. Grounded theory allows the researcher to begin with the question, collect data, examine ideas and concepts, extract and categorize that data to use it to form the basis of a new theory. This new theory can then be applied and tested statistically (Akbıyık, 2012). To successfully accomplish this, the approach for the study was fragmented into a three-part mixed methods study. A qualitative section utilizing semi-structured interviews and Delphi study with experts in the field followed by a quantitative section to test relationships between core concepts derived from the qualitative section. The qualitative portion of the study was done first, which allowed relationships to be tested later in a quantitative manner using statistical techniques. The knowledge gained through such a process allowed the quantitative section to be further insightful, concentrated and exploratory in nature.

Furthermore, it is important to note that this study also examines relations (Kerlinger and Lee, 2000) as it is being conducted to determine the relationships for successful Big Data projects. Standard strength and direction of relationships between variables are examined and predictions provided given the strength and conclusive nature of the variables within the study. The step by step process to investigate the research problem is as follows:

1. The first step was to be formally educated on both these topics. As the researcher was already on the journey to obtain a Ph.D. in Management Information Systems it was vital to enhance knowledge on Big Data. From a professional standpoint, the researcher works with a consultant on Big Data projects. Even so, education was needed to familiarize with various tools and techniques that professionals use in this trade every day. The researcher started by speaking to multiple global, startup and mid-size organizations, joined related LinkedIn professional discussions groups and took up reading the latest on Big Data. All this was done to increase the knowledge and skill level with the goal of being able to conduct semi-structured interviews and Delphi study and have conversations with professionals. This was an evolving process started in June 2017.
2. The second step was conducting semi-structured interviews with experts about what does “success” mean in Big Data projects. The research model consists of the CSF variables and “success” as the dependent variable. Delphi and computer-assisted telephone interviewing (CATI) rounds are utilized to form the CSFs. The semi-structured interviews with experts aim to enlighten what “success” meant for a Big Data project. The analysis of the semi-structured interviews generated the keywords and so the scale for success variable.
3. The next step was to start conducting Delphi study with professionals who have worked on and implemented Big Data projects, programs, and solutions. This was the qualitative phase. The researcher utilized their personal and professional network to find professionals and organizations who had implemented Big Data initiatives, solutions, projects and/or programs and who were willing to speak about their experiences which is commonly referred to as the “purposeful sampling” technique in qualitative research. The purpose of the Delphi study was to get feedback about success factors of Big Data projects. The initial goal was to speak with roughly 10-20 professionals regardless of industry, profession or location.
4. After conducting the Delphi Study, the researcher would look for common success factors that can be grouped into concepts and further into categories of CSF that can be measured as variables using a survey.

5. The final step was the creation of the survey. This was the quantitative (survey) portion of the mixed methods study. This would then allow the researcher to run statistical procedures to determine various CSF for Big Data projects.

The integration of the qualitative and quantitative design for this research allowed the researcher to help better understand, compile and relate Big Data Projects with critical success factors. This integration, as Creswell and Plano (2011) elude to allow for a single study to provide a more complete analysis of the research question being explored. In other words, we take one set of data, perform analysis and apply our learnings to build the other data set. This helps to further expand on the learnings gleaned from just the primary method. As such, mix method allowed the researcher to holistically look at factors impacting successful Big Data Projects. As we will review here, the qualitative research was conducted prior to the quantitative study. The learnings gathered from the initial qualitative analysis allowed the researcher to create a scale to statistically analyze the hypotheses and the quantitative research question.

Consequently, a sequential exploratory mixed methods design was selected where a qualitative phase informed a quantitative phase. The sequential nature of the research design was shown in Figure 7 below.

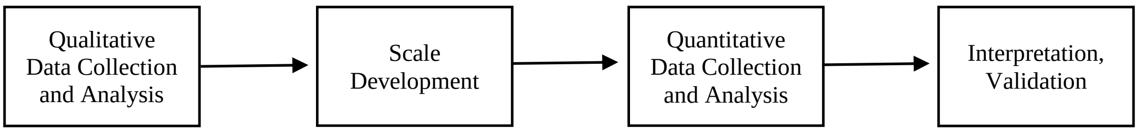

Figure 7: Sequential Exploratory Design

This mixed methods study design was optimal for developing a survey that required being informed through access to key participants in the field who ultimately represented the audience of interest. Semi-structured interviews and Delphi study were conducted to clarify and bound the CSF concepts and issues to explore. This initial qualitative study helped us gain an understanding of the situation related to Big Data projects and CSFs; particularly issues that emerged from the point of view of Big Data experts not readily identifiable through the published literature. The findings from the semi-structured interviews, Delphi study, and an intensive literature search were used to inform the instrument (a survey questionnaire) sent to Big data professionals.

Table 2: Methodological Descriptions

ITEMS	DESCRIPTIONS
Nature of study	Sequential exploratory
Sample Unit	Employees
Sample Criteria	Employees who have Big Data project experience
Sample frame	IT workers directory
Geographic scope	All regions of Turkey
Sampling method	Purposeful sampling
Method of collection	CATI (computer-assisted telephone interview)

This was highlighted in the figure below, wherein the exploratory qualitative phase (induction) of the design helped formulate grounded theories of CSFs of Big Data Projects, and the confirmatory quantitative phase (deduction) helped test and finalize the scale.

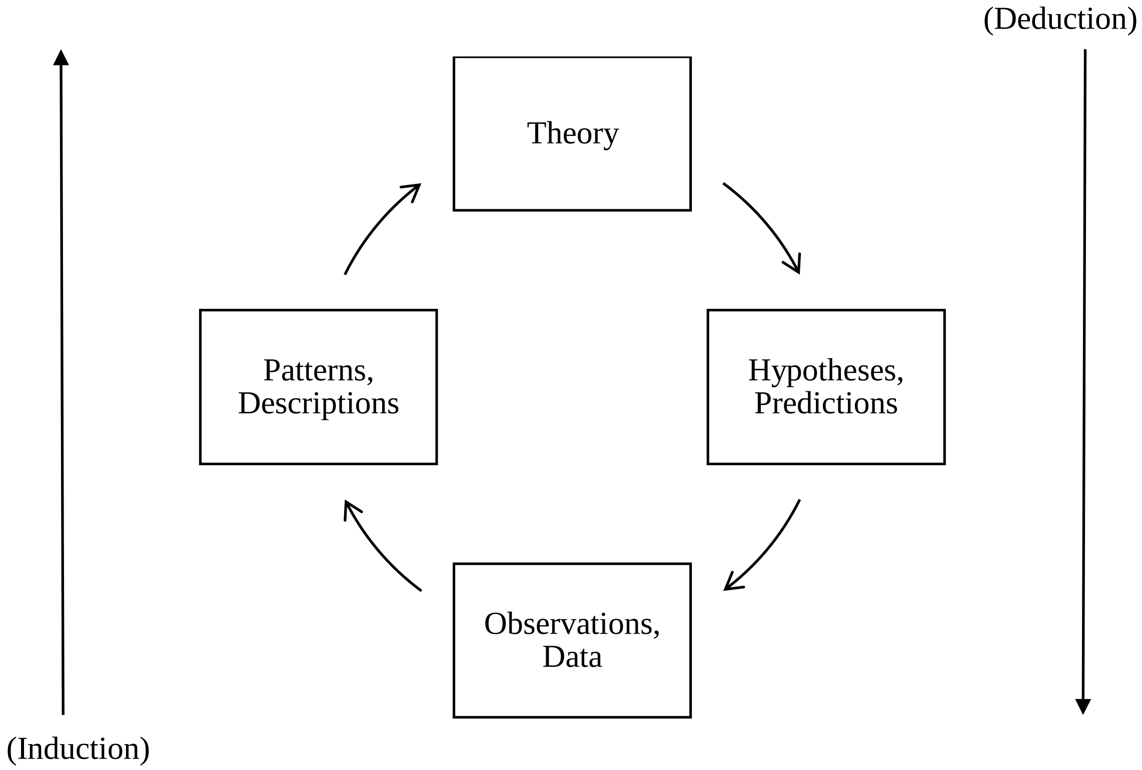

Figure 8: The research wheel

Source: Johnson and Christensen (2012)

3.3. Systematic Literature Review

The purpose of the literature review conducted for this research is to examine the previously done research within the field of Big Data projects and project success issues and to identify evidence related to CSF criterion. This will add to the body of Big Data project knowledge and be useful for developing suggestions for areas of further research. The focus of this research study is on the successful completion of Big Data projects. A deeper understanding of the literature will aid in the definition of CSFs, Big Data projects, project management, and success. These definitions and concepts will be used throughout the following research to answer the research questions: "*What are the critical success factors that impact project success in Big Data projects*? and "*What are the relationships among the critical success factors?"*. The materials available through the Sakarya University library were utilized, with the narrowed criterion of work created from 2009-present, peer-reviewed, and content that contains the full text. The results from these searches had to fall within these six criteria: availability within databases and journals, containing the full article, be peer-reviewed within the journals, and contain relevant information about Big Data or CSFs as it pertained to organizational structures, human structures, technology infrastructure, project management, cost and schedule management, and or leadership skills etc. Search terms used included: "big data" + project, "big data" + business, "big data" + success, "business intelligence" + project, "business intelligence" + success.

For a comprehensive literature review, this research utilized a custom structure borrowed from Creswell, (2009) and Cornell University (2016). The materials reviewed included books and journal articles. The databases used to search for research materials included: Web of Science and Scopus. After the systematical literature review, other databases (Elsevier ScienceDirect, Wiley Online Library, Elsevier ScienceDirect, Wiley Online Library, Sage and Springer, South Western, Oxford, Emerald Insight, IEEE, JSTOR and Springer) are also included for further details regarding specific issues. The framework for the systematic literature review used in this research employed seven steps 1) Identify the research question(s); 2) Define inclusion and exclusion criteria; 3) Search for studies; 4) Select studies for inclusion based on pre-defined criteria; 5) Extract data from included studies; 6) Evaluate the risk of bias of included studies; 7) Present results and assess the quality of evidence" (Creswell, 2009; Cornell University, 2016).

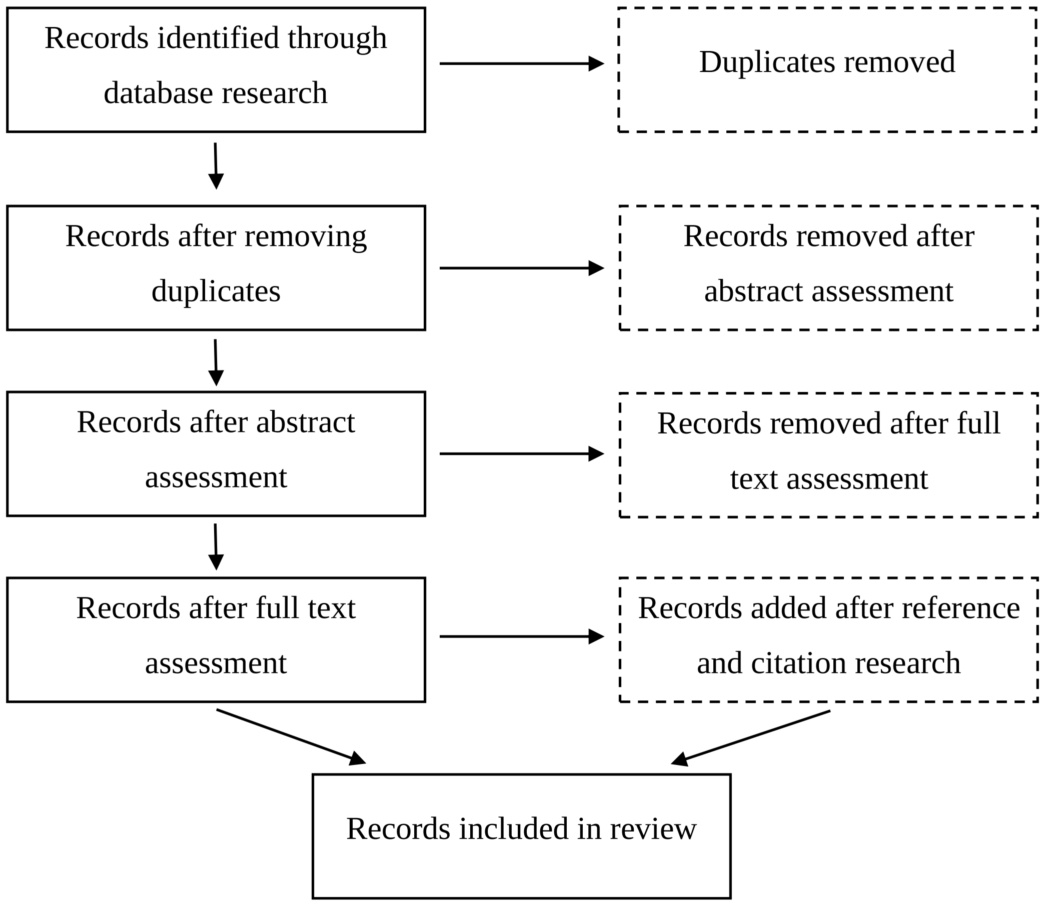

Figure 9: Systematic Literature Review Process

In existing literature, Big Data appears ~ 30,000 times across Web of Science core databases. At this juncture, the research utilized only peer-reviewed/scholarly/academic journals that were to be most commonly used by academics and practitioners alike for acquiring information and disseminating new findings and represent the highest level of research (Wamba et al., 2015 quoted from Niagi and Wat 2002). That bought the search down to 11,877 articles without any year limitations.

The purpose of this literature review was to address a deeper understanding of CSFs of Big Data projects which have come into question (Koskela and Howell, 2002; Mir and Pinnington, 2014). Since the focus was on non-technical articles but still considering the IS (information systems) side of things, the researcher wanted to further confine this result set with a term that incorporated many of Big Data findings and relevance to the organizations today. The research questions have required an analysis of their individual topics, from their origin through their evolution, and to the current practices and research findings. The research utilized various searches using terms such as management, organizations, marketing, analytics and information technology to name a few. The five terms that gave the best results and covered a large breadth of the Big Data CSF related landscape were, Big Data, project, success, business and business intelligence. Thus, limiting the search with the combination of these subject terms dropped the number down to 5445 main research articles spanning between the years of 2009 to 2018. Any further chopping or restriction removed certain articles and after the full-text assessment, 529 articles are examined as they covered a large surface area regarding Big Data. 10 records are added after reference and citation search and the systematic literature review is conducted on 539 articles.

Table 3: Systematic Literature Review Source Statistics

	WoS	Scopus	Records after full-text assessment	Records added after reference and citation search	TOTAL
«big data» + project	736	804	202	5	207
«big data» + business	926	1895	174	0	174
«big data» + success	286	301	83	2	85
«business intelligence» + project	101	146	34	2	36
«business intelligence» + success	122	128	36	1	37
TOTAL	**2171**	**3274**	**529**	**10**	**539**

Going through the literature, there were various articles on implementations, best practices, case studies, management/organization theories, complementing technologies, big data challenges, big data analytics and other multiple variations with each either providing proof (by theory) or via identifying challenges regarding Big Data. What was clear was that there was consensus that Big Data was deemed as the future, the real deal and central in creating big impacts (Halaweh and Massry 2015; Wamba et al., 2015;

Wixom et al., 2014; Xu et al 2015; Chen et al 2012; Forrester 2012; Church and Dutta 2013, McAfee and Brynjolfsson 2013; Manyika et al. 2013). Furthermore, there was no consensus on the definition of the term, Big Data (Hartmann et al. 2014; Young 2014; George et al. 2014; Church and Dutta 2013; Manyika 2013; McAfee and Brynjolfsson 2013, Wamba et al., 2015; Halaweh and Massry 2015), evidence of what influences successful Big Data projects.

A systematic review of the literature delineated details allowing a deep understanding of each topic. This provided the groundwork for answering the research questions, drawing conclusions and making recommendations for future research.

A major part of this systematic literature review process is summarized in chapter 1 and chapter 2, where definitions are reviewed and literature on factors that can impact successful Big Data projects are examined.

3.4. Mixed Methods Research

Mixed methods research is a methodology for conducting research that involves collecting, analyzing, and integrating (or mixing) quantitative and qualitative research (and data) in a single study or a longitudinal program of inquiry (Creswell, 2009). Methodologist John Creswell suggested a systematic framework for approaching mixed methods research. His framework involves four decisions to consider (Creswell, 2009, p. 211).

1. What is the implementation sequence of data collection?
2. What method takes priority during data collection and analysis?
3. What does the integration stage of finding involve?
4. Will a theoretical perspective be used?

Mixed methods provide a perspective from both the qualitative and quantitative approaches. Creswell also suggested six types of mixed methods designs that are likely to address most research inquiry situations:

1. Sequential exploratory: Qualitative phase informing a quantitative phase
2. Sequential explanatory: Quantitative phase informing a qualitative phase
3. Sequential transformative: Either phases conducted one after the other, with results integrated into the final analysis

4. Concurrent triangulation: Two methods used to cross-validate, confirm within a study
5. Embedded design: It gives priority to one approach which guides the study, while another is embedded
6. Concurrent transformative: The use of a theoretical perspective reflected in the purpose or research questions of the study to guide all methodological choices.

This research uses a sequential exploratory mixed method design in which the qualitative data set is embedded within and plays a supportive role in a study primarily based on the quantitative data (Creswell and Plano-Clark, 2007). The embedded design is used most often with experimental trials and is applied here as it best fits the research questions and the principles guiding this type of design: a single data set is insufficient to address the research issues, different questions are being investigated, and each question requires different types of data. The key elements of any mixed methods design include analytic logic, priority, timing of data collection and point of interface or mixing. Analytic logic centers on whether the two data sets are merged into one interpretation or analysis to address research questions or one data set are used to build on the results of a preliminary data set (Creswell et al, 2011).

The embedded design of the mixed method has shaped the parameters of the quantitative PLS-SEM (Greene, 2007). Therefore, although the qualitative and quantitative approaches are not intentionally mixed until the level of data collection, given the nature of the embedded design there is an interaction between the three components during the design process (Plano-Clark et al, 2013). As noted in other studies based on an experimental embedded model (Caracelli and Greene 1997, Plano-Clark et al, 2013), within this study, the design of the qualitative methods have been shaped by the requirements of the dominant PLS-SEM in the following major aspects: framing of research questions, sampling, level of control, data source, data collection procedures, and interactions between researchers and participants.

3.5. Method Appropriateness

The study adopts the positivist philosophy which postulates that knowledge is derived exclusively from information resulting from the interactions of logical and experimental methods. Positivism (Arghode, 2012) assumes that real-world situation exists, hence scientific methods can be used to explain the relationship between variables. The theoretical framework of this research defines the relationship between the independent variables and the dependent variable and explains how the research instruments operationalized the variables.

3.5.1. Qualitative Method Appropriateness

The qualitative method was considered as the first step of the quantitative approach to accomplish the study's objectives. The qualitative method is designed to gather participants' perspectives to create new theories. It is appropriate for research objectives related to gathering participants' various perspectives and then formulating theories based on data gathered (Creswell, 2005). The qualitative method has been sufficiently explored pertaining to the current research study topic.

3.5.1.1. Constructivist Grounded Theory

The most effective and interesting pursuit of the qualitative research question was determined by the researcher through a grounded theory methodology. The belief that the grounded theory methodology is acutely suitable for this qualitative research question was based on the methodology's efficacy for the construction of theory based on study data. Grounded theory methodology has several major schools of thought and each share many commonalities. This research leaned heavily on the systematic Straussian procedure for grounded theory methodology set forth by Corbin and Strauss (2015), in which categories, codes, and a constant comparative method of data analysis are used to develop emerging theories (Corbin and Strauss, 2015; Creswell and Poth, 2017).

As Strauss and Corbin state: "Grounded theory methodology incorporates the assumption, shared with other, but not all, social science positions concerning the human status of actors whom we study. They have perspectives on and interpretations of their own and other actors' actions" (1994:280). The grounded theory provides a fluid framework through which to inductively provide a conceptual interpretation and theoretical development of critical success factors for Big Data projects and the effects of these factors on project success. Grounded theory is applied as a methodology that develops a substantive theory of how teams or enterprises conclude Big Data projects with success. Strauss and Corbin argue that "grounded theory is a general methodology for developing theory that is grounded in data systematically gathered and analyzed" (1994:273). Theoretical frameworks of pragmatism and symbolic interactionism provide the conceptual scaffolding to uphold grounded theory approaches (Corbin and Strauss, 1990). These frameworks direct a non-determinist, fluid movement of data and the interpretation and reaction to the data by participants (Corbin and Strauss, 1990). This inductive approach to social science research emphasizes a constant comparative, iterative analysis of data collected throughout the project. This iterative process develops theory from the data. Originating in sociology, grounded theory is used to address research questions in the medical profession, anthropology, and other social science fields. Since the 1960s sociologists have discussed the evolution of the rules, procedures, and application of grounded theory.

In 1967, Barney Glaser and Anselm Strauss co-authored The Discovery of Grounded Theory to provide an in-depth guide to the rules and procedures specific to grounded theory application. Glaser and Strauss's (1967) co-authorship of grounded theory is now recognized as "Classic grounded theory" or CGT (Evans 2013). CGT theorists approach social science phenomena with no questions to specifically address at the beginning of research project. This allows for concepts to be discovered throughout the study. The literature review is adapted once emergent themes become more prominent throughout the study. These themes are arrived at through the use of substantive coding, theoretical coding, constant comparison, and memoing (Evans, 2013). The grounded theory developed by Glaser and Strauss is comprised of original rules, procedures, and application of grounded theory (Evans, 2013). Through its evolution, different interpretations of the grounded theory have developed. As Glaser continued to instruct classic grounded theory, Anselm Strauss redefined his procedures with Juliet Corbin in their 1990 book Basics of Qualitative Research: Grounded Theory Procedures and Techniques. Straussian-grounded theory redefined coding procedures and data structure through the development of eleven procedures (Corbin and Strauss, 1990; Evans 2013):

1. Data collection and analysis are interrelated processes.
2. Concepts are the basic units of analysis.
3. Categories must be developed and related.
4. Sampling in grounded theory proceeds on theoretical grounds.
5. Analysis makes use of constant comparisons.

6. Patterns and variations must be accounted for.
7. Process must be built into theory.
8. Writing theoretical memos is an integral part of doing grounded theory.
9. Hypotheses about relationships among categories are developed and verified as much as possible during the research process.
10. A grounded theorist need not work alone.
11. Broader structural conditions must be brought into the analysis; however microscopic in focus is the research.

Straussian coding is divided into three stages: open coding, axial coding, and selective coding (Evans, 2013). This is opposed to the two-stage classic grounded theory coding of substantive coding followed by theoretical coding. Straussian grounded theory is argued to provide a more rigorous and less fluid approach to grounded theory than that of CGT (Evans, 2013; Pandit, 1996). Classic grounded theory and Straussian grounded theory led to variations in the grounded theory approach. In order to address my project from a grounded theory perspective, a variation of this original version of grounded theory was applied. Constructivist grounded theory will be discussed later in this section. Feminist grounded theorists argue that grounded theory and feminist epistemology strongly align with one another in the understanding of experiences from the participants' perspectives. This perspective strongly aligns itself with a postmodern stance on questioning the truth of reality and our interpretations of that reality (Wuest, 1995). Feminist grounded theory is most often applied when researching female-specific roles and professional positions in a grounded theory framework. The continual evolution of grounded theory appears in the form of constructivist grounded theory.

Constructivist grounded theory challenges classic grounded theory by placing its focus on the construction rather than the discovery of concepts in grounded theory exploration (Evans, 2013). As opposed to CGT, constructivist theorists use literature to gain greater knowledge about what questions have and have not been addressed in a researcher's area of interest. When the literature review is implemented creates the greatest differentiation between classic grounded theory and constructivist grounded theory. Classic grounded theory argues that literature should not be reviewed at the beginning of a project for fear of skewing the analysis of emergent themes during research. Constructivists encourage review of the literature to gain a greater understanding of what topics have been researched in previous projects. Straussian and constructivist grounded theory share in their use of a literature review to provide knowledge of questions that have been addressed in the literature for the phenomena under investigation.

3.5.1.2. Delphi Technique

The purpose of conducting a qualitative study, utilizing a Delphi technique, is to identify the problems and CSFs associated with Big Data projects. The researcher for this study seeks to identify recommendations for improving these issues with Big Data. A qualitative method is suitable for the current research because the study is designed to investigate perceptions, experiences, and ideas (Ashby et al., 2015; Merriam, 2014). The qualitative technique is also useful for gathering a consensus opinion not found in the literature, an effort that would not be feasible with only quantitative approaches (Rees, Rapport, and Snooks, 2015).

The researcher will investigate perceptions, experiences, and ideas (Ashby et al., 2015). The researcher in this study has solicited and documented the opinions of subject matter experts and through this gathered information was able to ascertain a consensus on the subject of CSFs of Big Data (Kache, and Seuring, 2017).

The choice of the Delphi technique for this study is in preference to an interview-based design such as with a case study design. The Delphi technique is a collaborating method that can successfully provide innovative responses to researchers to gain answers to questions based on the design of the research model in use (Skinner et al., 2015). In a Delphi technique study, a panel of subject-matter experts receives inquiries regarding a specific subject (Guzys et al., 2015). The panel of anonymous experts proceeds with a prearranged number of questionnaire rounds with the goal of reaching a consensus on the subject (Rodriguez-Mañas, et al., 2012). Thus, the Delphi technique is the ideal tool of choice for increasing the overall understanding of multifaceted problems and for work in which different solutions are needed (Skinner et al., 2015).

In contrast, a case study is a thorough analysis of an entity, single event, or person (Yin, 2011). A case study design is not suitable for the current study because the purpose is to address input and recommendations from a group of Big Data experts and not to focus on a specific set of scenarios or organizations (Thomas, Silverman, and Nelson, 2015). We try to obtain a consensus from multiple participants after several rounds of questions, with the goal of defining the limitations affecting performance, or change. The Delphi technique is the design of choice for obtaining such a consensus from a group of subject-matter experts (Thomas, Silverman, and Nelson, 2015).

3.5.2. Quantitative Method Appropriateness

The quantitative method appropriateness pertained to the scientific approach for objectively collecting closely characterized numerical data to apply statistical formulas to produce information (Creswell, 2005). A quantitative method includes a process to examine the connection between variables by a statistical process to measure association in degrees and usability in predicting the outcome (Creswell; Devlin, 2006; Dyer, 2006). The current research questions pertained to establishing associations between predictor variables and the criterion variable. Statistical analysis was used to quantify the relationships between the 5 critical success factors and Big Data project success, with all variables measured on a 7-point Likert-type scale (Likert, 1932). The Likert-type scale is used in survey instruments by instructing participants to respond with a level of agreement to a declarative statement using a number scale (DeVellis, 2003).

The quantitative method was appropriate because the purpose of the study was to examine relationships among predictor variables with criterion variables and the effect on relationships. The literature was lacking empirical data to support relationships between the 5 critical success factors and Big Data project success.

This study used SAS CALIS procedure (SAS Institute Inc., 1990) to analyze the data with the maximum likelihood algorithm. Following a two-stage approach recommended by Anderson and Gerbing (1988), a measurement model was first estimated using confirmatory factor analysis to develop an acceptable latent structure, and then a structural model which specified the hypothesized causal relationships between the latent constructs were developed and assessed.

There are two types of causal modeling techniques: path analysis and SEM (Mertler and Vannatta, 2005). Path analysis takes into account the observed variables only and establishes a causal flow. Both causal direct and indirect effects can be estimated using path analysis. SEM takes into account observed and latent variables and offers many advantages over path analysis. It represents a melding of factor analysis and path analysis into a comprehensive statistical methodology (Kaplan, 2008). SEM is a confirmatory technique often used to test a theory; therefore, prior knowledge of theory or hypotheses about potential relationships among variables is required (Tabachnick and Fidell, 2007). A major advantage of SEM is that the computer analysis procedure provides an overall indication of the fit between the model and the theory, whereas such indication in path analysis is a manual process.

There are two types of Structural Equation Modeling (SEM). AMOS (Analysis of Moments Structures) software is the most commonly used method and utilizes covariance-based structural equation modeling (CB-SEM). The second type is SMART PLS (Partial Least Square), which performs variance-based structural equation modeling (PLS-SEM). Though CB-SEM is the more popular method; PLS-SEM is newer and the number of papers using PLS-SEM is increasing in many fields such as marketing, strategic management, management information systems, operations management and accounting (Hair et al., 2014).

3.5.2.1. Structural Equation Modeling

The main objective of employing SMART PLS software as a means of structural equation modeling is to maximize the explained variance of the endogenous latent constructs, known as dependent variables. SMART PLS is a latent variable modeling technique that incorporates multiple dependent constructs and explicitly recognizes measurement error (Karim, 2009). PLS-SEM is primarily used for theory development and exploratory research purposes when the research area is still relatively new or changing. CB-SEM aims for confirmation of a theory by determining how well a model can estimate a covariance matrix for the sample data (Hair et al., 2014).

Table 4: PLS-SEM vs CB-SEM Comparison

Model Requirement	PLS-SEM	CB-SEM
Includes interaction effects	Preferable, as it is designed for easy interactions	Difficult with small models, nearly impossible with large ones
Includes formative factors	Easier	Difficult
Includes multigroup moderators	Can use, but difficult	Preferable

Testing alternative models	Can use	Preferable, as it provides model fit statistics for comparison
Includes more than 40-50 variables	Preferable	Sometimes unreliable if it does converge; sometimes will not converge
Nonnormal distributions	Preferable (although it will still affect results, just to a lesser extent)	Should not be used; results in unreliable findings
Nonhomogeneity of variance	Preferable (although it will still affect results, just to a lesser extent)	Should not be used; results in unreliable findings
Small sample size	It will run (although it will still affect results negatively)	Unreliable if it does converge; often will not converge

Source: Lowry and Gaskin, 2014

There are less demanding conditions for sample size, independence, and normality imposed by PLS (Hair et al., 2013). Indeed, a study by Reinartz et al., (2009) shows that PLS requires only about half as many observations to reach a given level of statistical power as does CB-SEM when it comes to prediction and theory development.

A sample size of at least 200 is proposed by Hoelter (1983) in order to make an accurate assessment of model fit when using CB-SEM. PLS, on the other hand, is generally workable with smaller sample sizes (Gefen, Straub, and Rigdon, 2011) and when the assumption of normality is in doubt. This is due to the fact that PLS uses the original sample to estimate the model's parameters as it uses the re-sampling method (bootstrapping function) to calculate the confidence interval of the model parameters. Running the data using PLS seems to be a better choice rather than with CB-SEM.

Apart from that, PLS is able to handle both formative and reflective variables (Bollen, 2011) and has an advantage over a new investigation or study area where measurement items are newly developed (Anderson and Gerbing, 1988). This is supported by Chin and Newsted (1999) in that the PLS approach is more suitable when the phenomenon under research is relatively new or changing or when the theoretical measures are not well-formed. This is important in this study as the measurement items were extracted from the qualitative part of this study while some sentences were formed based on keywords provided in Delphi rounds.

In a meta-analysis of PLS-SEM review studies, the top three reasons for using SMART PLS are the flexibility on non-normal data, small sample size and the involvement of formative indicators (Hair et al., 2014). The advantages of employing SMART PLS over CB-SEM structural equation modeling can be listed as it has fewer restrictions on the sample size, it is not required for normal-distributed input data, able to analyze complex model with a multitude number of constructs, able to manage reflective and formative model with ease and finally when the purpose is to maximize the variance explained of the endogenous (Urbach and Ahleman, 2010).

In conclusion, PLS definitely has an advantage as it can explicitly recognize measurement errors while in AMOS errors need to be represented. This was of great help and convenience to the researcher when carrying out the data analysis. For this research, new variables are examined in a brand new research model and measurements of constructs are shaped in the light of Delphi results. In addition, the model involves reflective and formative constructs. Given all these reasons, it is deemed to be more appropriate to choose SMART PLS over CB- SEM based software.

Accordingly, a measurement model was first developed and assessed using confirmatory factor analysis (CFA) to ensure that the variables extracted to reflect the same latent factors were indeed highly correlated with each other and therefore reliable. The two-step approach was employed by assessing the measurement and structural model. Overall, the purpose of model validation is to determine whether both the measurement and structural model fulfill the quality criteria for empirical work (Ringle et al., 2015; Urbach and Ahlemann, 2010). The following sub-sections discuss the guidelines used in this study to assess both reflective and formative measurement and the structural model of this study.

3.6. Ethical Considerations

The risk of harm associated with social and behavioral sciences research is considered present when human subjects are involved (Hoser and Nitschke, 2010; White, 2009). The identification, assessment, and remedy of such risks are informed by a growing body of research literature (Hoser and Nitschke, 2010; White, 2009; Caulfield, Rachul and

Zarzeczny, 2012) on risks with human participants in a research study. The current research used CATI data on critical success factors of Big Data projects, which involved human subjects. For the semi-structured interview section, there were absolutely no recorded conversations to maintain the confidentiality of the interviews. CATI data is recorded without any personal or contact information. The sampling procedures that were used in this study involved a random sample drawn from Big Data professionals. No identifying information was collected via the survey as the results were completely anonymous. All of the data was aggregated so that no individual responses could be identified. There were minimal risks to participants. This study followed the ethical principles found throughout the Belmont Report, which provides ethical principles and guidelines for studies involving human subjects. All participants were treated as autonomous agents. Since the study used a random sample, the responses will automatically be anonymous. Responses were only used in aggregate form and no identifying information was gathered in the survey, such as names or addresses. The survey does not include any personal questions regarding individuals' privacy and confidentiality according to the privacy act.

The principle of beneficence was also incorporated into the study so that no harm would come to anyone participating in this study. In fact, participants may benefit by receiving a copy of the results, which could assist them in improving the success of data mining projects. The survey took participants approximately 15 minutes to complete, thus, there was very little risk to participants. The principle of justice was also present in this study because every member of the sample had an equal opportunity to participate and was given the option to opt-out at any time. All data collected as part of this study were downloaded and archived in the researcher's safe. This archive will be discarded after four years. Electronic copies of the data will be discarded in a secure manner.

Ethical assurances were taken to protect participants from harm and to safeguard anonymity and confidentiality. Ethics committee approval is included in the Appendix.

CHAPTER 4: DETERMINING CRITICAL SUCCESS FACTORS

The analysis began with simple coding methods for emerging common themes. The codes were created specifically for the study. Due to the uniqueness and exploratory nature of the study, generalized coding systems were not implemented. Through semi-structured, one-on-one semi-structured interviews and Delphi study codes were developed and further analyzed through the transcribed notes by the researcher. From this point, the analysis of the written data by the researcher was used to develop codes and determine emerging themes.

Twenty-five codes emerged during the study. In order to analyze these emergent codes, each question addressed in the thesis was used to provide a code grouping framework. The following three questions were addressed in the semi-structured interviews*: "What do you mean by a successful Big Data project?"*, *"How do you know the project was successfully completed."* and *"How do you define a successful project?".* In the Delphi study participants were requested to answer the following two questions: *"What are the critical success factors for successful completion of Big Data projects?"* and "*In case of lack of which factors it becomes difficult to complete a Big Data project successfully?".* Codes have been reviewed and applied to answer the qualitative research question addressed in the study.

4.1. Research Timeline

The timeline for this qualitative research was from June 2017 to August 2017. Semi-structured interviews started in early June and took 2 weeks. In early July, Delphi questions were shared with the committee for fine-tuning. In late July, the researcher gathered the first round answers from experts. The second round of the Delphi study is continued in August.

1.2. Semi-Structured Interviews

As described in the previous chapters, before the content of a new scale can be drafted, the researcher must define and understand the underlying construct, and articulate its connection to relevant existing theories, to aid to clarity in scale development (Clark and Watson, 1995; DeVellis, 2016). Chapter 2 described the related literature and existing theories that inform an emerging theory of CSFs of Big Data projects. Existing related theories help identify the boundaries of an emerging phenomenon so that the scale does not unintentionally drift into other domains (DeVellis, 2016). Theories can drive, and also be the outcome of, the research process of coding qualitative data (Saldana, 2009). This study, as previously discussed, began theoretically, with qualitative data that will directly lead to scale items; Big Data Success Scale can then in consequent research be used to develop theory and test hypotheses.

Therefore, the first step of devising the Big Data Success Scale is to formulate a definition of the phenomena of "Big Data project success" and describe how this construct relates to other phenomena and their operationalization (DeVellis, 2016). A database of well-organized raw data forms a chain of evidence that allows the researcher to demonstrate that her interpretation of the data is firmly grounded in the data (Lazar, Feng, and Hochheiser, 2010; Yin, 2003). In the present research, such a database starts with the results of a semi-structured interview study of the intended population. The definition of the "success" construct (also the dependent variable) then emerges from these data.

In the development of a scale such as the Big Data Success Scale, an item development study as an initial phase of scale development can form the basis of an argument for content validity.

Interviews can help the researcher understand the thinking and the vocabulary of the target group, and discover topics addressed by potential respondents (DeWalt, Rithrock, Yount and Stone, 2007). DeVellis (2016, pp. 60-61) described the process used by Sterba, DeVellis, Lewis, Baucom, Jordan, and DeVellis (2007) to form the basis for content validity: "The study aimed at identifying appropriate content from the broader empirical and theoretical literature for possible inclusion in the measure. Although the authors examined content from measures of related constructs they geared their item development to specific features of the construct as they [participants] had defined it."

1.2.1. Method

Before the Delphi study, semi-structured interviews with experts were conducted via phone (6 out of 17) and in most cases, face to face (11 out of 17) due to the sensitivity and confidential competitive advantage information regarding Big Data. There were absolutely no recorded conversations to maintain the confidentiality of the semi-structured interviews (only the researcher and research committee are aware of specifics).

Some even went as far as to ensure legal compliance as well. The researcher provided each with reason and nature of the study and ensured strict confidentiality. The researcher also provided the ability for the interviewee to opt-out of the interview at any time and not share specific details if they did not feel they wanted to. The researcher did take notes around contextual and characteristic specific findings (such as technical skills required or involvement of teams, etc.) to create a survey to statistically establish and research other relationships.

According to Lazar, Feng, and Hochheiser (2010) and Yin (2003), interviews can be analyzed using various qualitative data analysis methods. Such methods identify common or repeated themes and structures among and within participants. One such technique, implemented in this study, is content analysis. The researcher conducts content analysis by examining the frequency of terms that may indicate concepts and the relationships among concepts. It assumes that the interviewee's comments evidence what he or she finds important, and why (Robson, 2002).

Another approach is to categorize interview content, which is either pre-defined or defined after analyzing the text (Lazar, Feng, and Hochheiser, 2010). In this case, categories were identified using the substance of the interviews. Ideally, interview results are presented with specificity and clarity, e.g., providing exact frequencies of a type of comment and using the interviewee's choice of words (Lazar, Feng, and Hochheiser, 2010). This method, as well, was implemented in the present analyses.

Interviews ranged in time from 30 to 80 minutes, with a mean length of 45 minutes. Each participant was interviewed about the term "project success" and the concept of "Big Data." The questions were formulated following the comprehensive review of the literature and were thus informed by its findings. Both the interview questions and the method of analysis for each is described in the next section. In this study, the researcher typed notes of the participants' responses during the interview, and there were absolutely no recorded conversations to maintain the confidentiality of the semi-structured interviews (only researcher and research committee are aware of specifics). After each interview, the researcher analyzed the notes and entered keywords and phrases in a spreadsheet, to collect frequencies on responses that could be quantified and to distill comments into common categories.

1.2.2. Sampling

For the Big Data Success Scale, the target population was defined as Big Data experts. Purposeful and snowball sampling was used to recruit the study participants. Purposeful and snowball sampling was adequate for this research because they allowed for the selection of highly probable information-rich big data professionals which best illuminated the research questions (Patton, 2015). Moreover, research on profession requires purposeful, deliberate sampling in order to accurately select participants. Purposeful sampling is an ideal way to represent the average person, situation, or instance of a particular phenomenon (Merriam, 1988). Participants for this study were chosen on the premise that they were able to provide a perspective on the phenomena under investigation (Smith, 2015). Due to the nature of the research, the respondents must be seasoned professionals, and they must be highly knowledgeable and skillful in the subject matter. Therefore, the study's sample was a purposeful sample of Big Data experts with a minimum of 5 years of experience in Big Data projects.

Interviews were conducted with the target population (experts) to gain information on this population's conceptualization of the construct "Big Data project success" and the language they use to talk about it. The objective of such interviewing in scale development is to use the resulting key phrases and ideas gathered from the target population in defining the construct and in writing the initial item pool (DeVellis, 2016; Clark and Watson, 1995).

The researcher reached out a total number of seventeen experts, two of whom were female. The number of participants used was more than other studies implementing interviews in scale development, e.g., Yildirim and Correia (2015) interviewed nine people in developing their nomophobia questionnaire. The ratio of male to female is not a limitation of the present study, and the researcher was not sensitive to this shortcoming. Because interviewing is resource-intensive, large representative samples are generally not possible; however, interviewing does result in a rich qualitative data set (DeWalt, Rothrock, Yount, and Stone, 2007).

The participants ranged in age from 30 to 47, with a mean age of 35. All participants had at least graduate degrees. Working industries of the experts included banking, pharmaceutical, energy, mobile communication, mobile application, technology, automotive, retail, cosmetics, health, building, telecommunication, fuel, food, garment, education, white appliances.

1.2.3. Results

Results from the interview questions are presented and discussed below. When appropriate, data tables are used to show how the participants' responses were categorized and quantified. Categories are presented in order of frequency of responses that were assigned to that category.

The following three questions were addressed in the semi-structured interviews:

- "What do you mean by a successful Big Data project? Start with the top 3 words or phrases that come to mind."
- "How do you know the project was successfully completed? Start with the top 3 words or phrases that come to mind."
- "How do you define a successful project? Start with the top 3 words or phrases that come to mind."

Answers were examined for conceptual commonalities and tallied. Six categories emerged from the data, as shown in Table 5.

Table 5: Semi-Structured Interview Results

Category	Phrases
Reaching project schedule	Finishing on time, delivery on time, be careful with the deadlines, reaching project schedule
Reaching project goals	Alignment with project purposes, reaching project goals, reaching project targets
Reaching quality goals	Reaching quality criterion, comply with the quality criterion, be careful with quality expectations
Reaching project budget targets	Compliance with the budget amount, to not exceed the budget, be careful with budget limits, reaching project budget targets
Satisfaction of end users	End-user satisfaction, shareholder satisfaction, compliance with end user expectations

Category	Phrases
Perception of success	To believe that the project ended with success, to feel that the project ended with success, to think that the project ended with success, the perception of success, to inherently know that the project ended with success

1.3. Delphi Study

Developed by Norman Dalkey and Olaf Helmer at the Rand Cooperation in the 1950s (Franklin and Hart, 2007; Hsu and Sandford, 2007; Mayfield, Wingenbach, and Chalmers, 2005), the Delphi technique was first used in technology forecasting for military use (Hanafin, 2004; Martin and Frick, 1998). The Delphi technique provides an organized method to gather perspectives from people with proficiency on a certain topic (Dalkey and Helmer, 1963). An advantage of the Delphi technique is that panelists are not required to gather for in-person discussions. Therefore, the proximity of the panelists is not a concern for researchers intending to employ the Delphi technique.

The Delphi technique is an effective method of group communication, allowing panelists with extensive knowledge on a certain topic to solve problems (Linstone and Turoff, 1975). The Delphi technique has been widely used in IS research in areas such as information systems management (Brancheau et al., 1996; Doke and Swanson, 1995), accounting systems (Worrell et al., 2013), health information systems (Hübner-Bloder and Ammenwerth, 2009; Snyder-Halpern, 2001). This study adopted the Delphi technique in order to reveal CSFs, as the Delphi technique promoted individual thinking while guiding participants toward consensus. Similarly, researchers have employed the Delphi technique to determine the CSFs regarding e-learning (Bhuasiri et al., 2012), m-commerce (Xu and Gutierrez, 2006), knowledge management (Yew Wong, 2005) and business intelligence (Yeoh and Koronios, 2010).

Three features of the Delphi method include anonymity, controlled feedback, and statistical group response (Dalkey, Rourke, Lewis, and Snyder, 1972). The Delphi technique aims to reach a consensus concerning a specific topic through rounds of questionnaires (Hanafin, 2004; Hsu and Sandford, 2007). The outcome of the three-round technique begins with the initial round generating a variety of answers, generally by asking panelists to answer one or two open-ended questions (Ludwig, 1997). Panelists provide information they believe will successfully address the question at hand (Linstone and Turoff, 1975). In the second round, panelists are asked to “review the items summarized by the investigators based on information provided in the first round (Hsu and Sandford, 2007, p. 2). As the second and third round follow, individual responses converge, resulting in a more accurate and defined group response of the initial question (Dalkey et al., 1972).

1.3.1. Process and Compilation

The Delphi technique is rooted in two traditional approaches: Conventional and Conference (Linstone and Turoff, 1975). The Conventional, or the paper-pencil, approach involves administering a questionnaire with a series of questions to the selected panel. The Delphi Conference approach utilizes computer technology to administer questionnaires and gather panelists’ responses (Linstone and Turoff, 1975). Stitt-Gohdes and Crews (2004) noted a benefit to the Delphi Conference is that it promotes faster response times as there is less delay in sending the rounds of questionnaires.

After the panelists provide answers to the solicited questions, a second questionnaire is developed based on their responses and administered to the same panel (Stitt-Gohdes and Crews, 2004). The rounds of questionnaires and feedback are continued until a consensus is met on the statements in question (Stitt-Gohdes and Crews, 2004).

A review of literature conducted by Martin and Frick (1998) found a majority of research studies employing the Delphi technique used modifications. Guided by Ramsey (2009), the present study used a modified Delphi technique of three rounds instead of the traditional four. According to Brooks (1979), Custer, Scarcella, and Stewart (1999), and Ludwig (1997), administering three rounds of questionnaires often is satisfactory to reach consensus among panelists. Using two panels of experts instead of one was another modification implemented by the researcher. “Using two panels allowed the researcher to compare the items that reached ‘consensus agreement’ within the two panels” (Ramsey, 2009, p. 54). Appropriately, a modified Delphi technique was used in this study.

The researcher reached out a total number of 17 experts including banking, pharmaceutical, energy, mobile communication, mobile application, technology, automotive, retail, cosmetics, health, building, telecommunication, fuel, food, garment, education, white appliances industries using their personal and professional network.

The researcher sent emails to potential panelists inviting them to serve on as experts in this study. Panelists who agreed to participate then received an additional email containing instructions for completing the first questionnaires and a hyperlink to the online instrument. The first round’s questionnaire initially was developed by the researcher in Microsoft Word 2016 and then transferred into Google Forms, an online surveying software. After collecting responses from the first questionnaires, the second round’s questionnaires were sent to panelists asking them to rank their level of agreement with CSF statements found in the first round. The experts reached agreement on the second panel and the researched concluded the rounds.

1.3.2. Validity

Ensuring face and content validity of the instruments used in the present study was a priority to the researcher. According to Creswell (2005), validity is concerned with assuring conclusions drawn from the instruments are accurate and represent what the instruments intended to measure. Privitera (2017) defined face validity as a judgment of which an instrument appears to measure what it intends to measure. Content validity determines whether the instrument can successfully represent and measure the construct in question (Privitera, 2017). Questionnaires for each round were examined for face and content validity by a panel of experts. This panel consisted of faculty members from the Sakarya University School of Business staff. The researcher consulted with the expert panel to enhance the validity of each questionnaire administered in this study. Expert panelists provided constructive feedback, suggesting minor revisions on the instruments before the researcher disseminated them to the participants. The researcher used the feedback to clarify the wording of the introduction and ensure there was uniformity in the scales in each instrument.

1.3.3. Reliability

Reliability of an instrument is determined by the consistency and stability of the constructs it measures (Creswell, 2005). Although no consensus regarding an optimal Delphi panel size exists in the literature (Hsu and Sandford, 2007), Dalkey et al. (1972)

reported an increase in reliability of group responses as the panel size increased. Yet, Sutphin and Camp (1990) stated panels should include an adequate number of participants to achieve intended results, but advised against including an overabundance of panelists as it results in excess data not beneficial to the study. A correlation coefficient of 0,9 was found with a group size of at least 13 panelists (Dalkey et al., 1972). To that end, 17 panelists remained in the final panels, solidifying the reliability of 0,9 outlined by Dalkey et al. (1972).

1.3.4. Sampling

Using the Delphi technique offers numerous benefits such as promoting strong participation from groups who are often left out of research (Brady, 2015). One advantage of the Delphi technique is that it acknowledges the unique contribution of each panelist (Hanafin, 2004). “The Delphi method is not concerned with having a generalizable sample but instead seeks input from a purposive sample of individuals with specific expertise on a topic” (Brady, 2015, para. 2). Panel selection is an important component of a successful Delphi study. Panel members must be knowledgeable on the subject in question (Brooks, 1979). Random selection is not an appropriate tool to generate a Delphi panel, and the researcher should carefully consider the knowledge of the potential participants and define the participants’ expertise, characteristics, and qualifications before identifying a sample from which to recruit (Brady, 2016; Ludwig, 1997). A differing trend from traditional focus groups is that panelists in a Delphi study remain anonymous to each other (Fletcher and Childon, 2014).

Purposeful and snowball sampling was used to recruit the study participants. Purposeful and snowball sampling was adequate for this research because they allowed for the selection of highly probable information-rich big data professionals which best illuminated the research questions (Patton, 2015). Moreover, research on profession requires purposeful, deliberate sampling in order to accurately select participants. Purposeful sampling is an ideal way to represent the average person, situation, or instance of a particular phenomenon (Merriam, 1988). Participants for this study were chosen on the premise that they were able to provide a perspective on the phenomena under investigation (Smith, 2015). Due to the nature of the research, the respondents must be seasoned professionals, and they must be highly knowledgeable and skillful in the subject matter. Therefore, the study's sample was a purposeful sample of Big Data experts with a minimum of 5 years of experience in Big Data projects. Therefore, purposeful sampling helps to ensure that the participants selected are able to articulate, express, and explain the phenomenon being explored in a clear and concise manner (Palinkas, Horwitz, Green, Wisdom, Duan, and Hoagwood, 2015). Moreover, studies reveal that purposeful sampling is critical when factors such as availability and willingness to participate are of concern to the researcher. Additionally, snowball sampling enabled the researcher to obtain additional participants via referrals made by individuals who were aware of others who share some or all of the characteristics that matched the particular area of research interest (Biernacki, and Waldorf, 1981). Much like purposeful sampling, snowball sampling is particularly appropriate when the subject matter is sensitive in nature and is one that requires knowledgeable insiders to assist in locating other participants.

The researcher had reached out to more than 30 individuals and not everyone responded. In the present study, 17 volunteer experts were recruited to serve on panels.

1.3.5. Data Collection

Using a series of questionnaires, the Delphi technique collects data from a selected panel in an attempt to build consensus (Dalkey and Helmer, 1963; Linstone and Turoff, 1975). This study sought to determine expert perceptions of the CSFs of Big Data projects. Throughout the duration of this study, both panels remained separate from each other and were administered instruments specific to each panel. Before each round, the researcher sent emails to panelists containing instructions for completion and hyperlinks to access each questionnaire. Questionnaires were administered and data collection was executed through Google Forms. Panelists were given one week to complete the questionnaires in each round. The researcher made the decision to eliminate panelists from the study who did not complete the instrument to which they were provided in rounds one and two. Details of procedures employed in each round of the study are described below.

1.3.6. Panel One

The first questionnaires were sent electronically to panelists serving on both panels on June 16, 2017. A reminder email was sent on June 23, 2017, to combat attrition of the panel sizes. The first questionnaire solicited personal and professional characteristics of the attendants. Such characteristics included sex, age, and position. Panel one included an open-ended question,

"What are the critical success factors that impact project success in Big Data projects?" Please do a personal brainstorm in the light of this question to identify and list as many factors as possible.

Each expert has listed and submitted his/her thoughts about the CSFs. Similar items in the responses were categorized by the researcher and the advisor and they are listed for rating in the second round.

Table 6: Delphi Panel One Results and Categorization

Phrases	Category
Money, investment in new technologies, human resource employment	**Financial efficiency**
Change management, managing problems related to change, overcoming problems about changes in business processes	**Change management**
Recruiting especially for the project, managing recruitment processes, hiring just for the project, finding right skilled employees	**Recruitment strategy**
The leadership of team leader, organizing skills of team leader, communication skills of team leader, skills management	**Team leader skills**
A multidisciplinary team, people from all departments, end-user collaboration from different departments	**Multidisciplinary team**
Team skills, analytic skills, technical skills, team competency, right people with right skills, communication skills, mathematical modeling, data visualization, statistical skills	**Team skills**
Education status of the team, educated analysts, education on technology, Big Data related education, Big Data related skills	**Education status of the team**
Team communication	

Phrases	Category
communication between team members and end users	**Communication ability**
Establishing appropriate infrastructure to meet the needs, sufficiency in the future, technological competence for a long time	**Technology infrastructure**
Big data strategy, hardware software tools of the new analysis platform, Hadoop, Python, Pig, Hive etc.	**Defining technology strategy**
Database management system, data eco-system, data must be organized, data control and administration; data quality	**Data quality**
To follow the latest technology, innovative analysis tools, keeping up with new technology advancements	**Keep up with technology trends**
Infrastructure, hardware, software, technical infrastructure must be suitable for the integration of data, integration of old and new databases, integration of data from different sources	**Technology infrastructure**
Easy access to data sources	**Easy access to data sources**
Task - technology - people balance, the appropriate employee for the new technology, adequate technology solutions for complex problems	**Task - technology - people balance**
Accurate deployment of time and resources	**Allocation of resources**
Proper job descriptions, managing team, arrangingthe the right job for the right person	**Team management**
Scheduling, effective progress, providing benefit in a short time, rapid results, immediate implementation of results	**Project schedule**
Achieving measurable outputs	**Defining measurement of success**

Phrases	Category
Documentation, saving the project process, team members should be careful about documentation throughout the project because the system will be constantly updated	**Documentation**
Correct definition of the business problem, correct definition of the project objective and scope, what are we aiming to add value to?, compatibility with business objectives	**Defining business needs and objectives**
Suitability, appropriateness, fitting the needs	**Suitability**
Strategic position of Big Data should be determined	**Positioning Big Data within enterprise**
Business structure, enterprise structure, decision-making process, how the business is carried out, the way they work, the importance of data usage in the workflow, the importance of Big Data and outputs, data-driven processes	**Compatibility with business processes**

Seventeen panelists completed panel one. Eighty-three statements were analyzed by the researcher, combining comparable comments and separating compound statements (Shinn et al., 2009). Through detailed thematic analysis, concepts and categories were developed, leading the researcher to identify 24 CSF statements representing CSFs of Big Data projects.

1.3.7. Panel Two

Panelists who completed round one were asked electronically to participate in round two. Round two questionnaires were generated based on the responses gathered from the first round and included 24 items identified at the first round by the Big Data experts and 1 item identified by the researcher from the literature. The second questionnaires were sent electronically on August 11, 2017, to the panelists who completed the first round: A reminder email was sent to the panelists who had not yet completed the second questionnaires on August 18, 2017.

Panelists were asked to rank their level of agreement with each CSF for Big Data projects. A seven-point Likert scale was used (Calisir and Calisir, 2004): 1=Strongly disagree, 2=Somewhat disagree, 3=Disagree, 4=Neutral, 5=Agree, 6=Somewhat agree, 7=Strongly agree. Harnessing the controlled feedback characteristic of the Delphi technique, summaries (categories extracted from the first round) of the first round's interactions were distributed to the panelists (Hsu and Sandford, 2007). To assist the panelists, documents containing anonymous responses from the open-ended question in round one were attached to the respective second questionnaires, providing "an opportunity for the experts to respond and revise their answer in light of the group members' previous responses" (Fletcher and Childon, 2014; Ludwig, 1997). Further, Ludwig (1997) stated utilizing a feedback process helps Delphi panelists become aware of the variety of opinions among the rest of the panel. Comment boxes were included alongside each item for panelists to request clarification or share additional thoughts regarding the CSF items (Ludwig, 1997).

1.3.8. Data Analysis

Several analytic approaches in the Delphi method exist, and adoption of each approach is determined by the objective of the study (Brady, 2015). Data were analyzed through Statistical Package for the Social Science (SPSS) v20, Panelists' personal and professional characteristics were examined using percentages and frequencies. In the second round, the frequency distribution value percentage approach was employed to determine the status of agreement on each challenged statement (Buriak and Shinn, 1989). Brady's (2015) thematic analysis process, advised by Bazeley (2009), served as the guiding framework of the qualitative analysis of the first round in the present study. Thematic analysis was used to develop reoccurring themes present in the qualitative portion of the first questionnaires. Utilizing the thematic analysis process, qualitative data were examined by identifying concepts and categories, which were then compiled into themes (Brady, 2015). Concepts closely reflect the original data provided by the panelists while the broader, more generalized categories present exceeding explanation of data (Brady, 2015). Brady (2015) noted researchers must possess extensive knowledge of literature regarding the subjects at hand to organize data into appropriate concepts and categories.

Throughout the Delphi process, it is crucial to ensure measures are taken to eliminate research bias (Ludwig, 1997). The thematic analysis relies on the researcher's interpretation of the data provided by the panelists. Inadvertently, the researcher may insert their own bias into the analysis of the panelists' responses (Brooks, 1979). To ensure correct and definite representation of the panelists' responses, a spreadsheet detailing the researcher's thematic analysis was distributed in the second round (Brady, 2016). This panelist-check process was used to solidify the accuracy of the researcher's analysis (Brady, 2016). Panelists were given the option to comment on or clarify the validity of the thematic analysis, however, none chose to do so through the duration of the study.

The statistical methods used to analyze the second round responses were median, percentage and the interquartile range (IQR) to establish levels of agreement. The IQR assists in understanding the spread of a set of numbers which are organized in ascending order. It is defined as the difference between the upper quartile (the highest 22%) and the lower quartile (the lowest 25%) of a data set. Gracht (2012) recommends the use of the median and interquartile range rather than the mean and standard deviation for the reason that mean is solely valid with interval or ratio data, whereas the Delphi technique utilizes ordinal scales whose intervals or ratio cannot be identified. This is backed by Argyrous (2005) who stresses that the calculation of the mean for ordinal data is not the correct procedure citing that in group judgments, outliers can skew the mean unrealistically. The debate on the use of the mean for ordinal data remains, but for this research, the median to measure central tendency and the IQR to measure dispersion for the median were used to evaluate consensus. Consistent with a study done by Gracht (2012), as a rule of thumb, an IQR of 1 or less is usually found to be a suitable consensus indicator for 4-7 unit Likert scales. Gracht (2012), cites that the IQR is frequently used in Delphi studies and is generally accepted as an objective and rigorous way of determining consensus.

For this study, an IQR of 1 or less was found to be a suitable consensus indicator. However, because the IQR method, through rigorous, lacked complexity in separating the degree of agreement (it only indicated that there was either agreement or not), frequency percentages were also utilized to identify the levels of agreement.

There is no universally agreed proportion for the Delphi Survey and the level used will depend on the size of the sample, the aim of the research and resources. Loughlin and Moore (1979) suggested 51% agreement amongst respondents, Sumsion (1998) recommends 70%, while Green et al. (1999) opted for 80%. More than 67% on a nominal scale or yes/no responses was considered consensus (Alexandrov et al., 1996 and

Pasukeviciute et al., 2001) while Putnam et al., (1995) opted for more than 80% on a 5-point Likert scale in the top 2 measures (desirable/highly desirable).

With reference to stopping at Round 2, in MacCarthy and Atthirawong (2003), it was assumed that another round would not significantly add to the results and therefore terminated the process. “Overall, it was felt that a third round of the study would not add to the understanding provided by the first two rounds and thus the study was concluded” (MacCarthy and Atthirawong, 2003).

For this study, the levels of consensus and qualifications for this research are summarized in Table 7.

Table 7: Levels of Consensus and Qualifications

Level of Agreement	Conditions
Consensus	IQR ≤1 and a percentage score ≥ 60% in a single level
Strong Agreement	IQR ≤1 and a percentage score ≥ 67% in combined adjacent levels
Disagreement	IQR > 1 and a percentage score ≥ 60% in a single level
	IQR > 1 and a percentage score ≥ 67% in combined adjacent levels
	IQR ≤ 1 but percentage score < 60% in a single level
	IQR ≤ 1 but percentage score but percentage score < 67% in combined adjacent levels
Total Disagreement	IQR > 1 and a percentage score < 60% on all scales and combined adjacent levels < 61%
Split Disagreement	Regardless of IQR, percentage scores > 25% on extreme ends

The interquartile range is the middle 50 percent of the ratings and lies between the first and third quartiles. An interquartile range greater than two indicates that the ratings were widely dispersed and the experts could not reach consensus. According to numerous studies using the Delphi Technique, including Heiko (2012), Passannante (1994), Becker and Roberts, 2009), and Basham (2010), an interquartile range of 2 or less indicates that consensus was reached. Therefore, for purposes of this study, the consensus was defined as any range that was two or less. Individual ratings that fell outside the interquartile range were flagged and expert panelists were given an opportunity to change their rating by moving their value closer to the median score. The interquartile range and the percentage

of expert responses which fell within the interquartile range for each identified element were calculated. These calculations are provided to illustrate the strength of consensus for each of the 25 elements for both importance and likelihood of implementation.
Items that had the defined levels of agreement (Consensus and Strong Agreement) were accepted as CSFs for Big Data projects (Ramsey, 2009; Shinn et al., 2009). The experts met consensus on all the 25 items, accordingly, the Delphi study concluded by the second round.

1.3.9. Results

The researcher's main intent for using the Delphi Technique was to reach a consensus among the experts. Semi-structured interviews and Delphi study are conducted on the same sample of experts. The researcher reached out a total number of seventeen experts. Of the seventeen experts who voluntarily gave consent to participate, all (response rate at 100%) returned fully completed first and second rounds of survey questionnaires.
The first round was a brainstorming round; the respondents could give their comments in a "comments section" provided the question. The open comments section of the questionnaire provided valuable feedback although this qualitative data was analyzed through thematic analysis. The intent of the first round of this study was to understand the perceived CSFs of Big Data projects by Big Data experts. Along with answering questions about their personal and professional characteristics, panelists responded to the open-ended question: "*What are the critical success factors that impact project success in Big Data projects*?". Seventeen panelists completed round one. Eighty-three statements were analyzed by the researcher, combining comparable comments and separating compound statements (Shinn et al., 2009). Through detailed thematic analysis, concepts and categories were developed, leading the researcher to identify 24 CSF statements representing CSFs of Big Data projects.

Table 8: Delphi IQR Results

NO	Category	Median	Q1	Q3	IQR
1	Financial efficiency	7	6	7	1
2	Change management	7	6	7	1
3	Recruitment strategy	6	6	7	1

NO	Category	Median	Q1	Q3	IQR
4	Top management support *	7	6	7	1
5	Team leader skills	7	6	7	1
6	Multidisciplinary team	6	6	7	1
7	Team skills	6	6	7	1
8	Education status of the team	6,5	6	7	1
9	Communication ability	6,5	6	7	1
10	Technology infrastructure	6	5,5	6,5	1
11	Defining technology strategy	6	6	7	1
12	Data quality	7	6	7	1
13	Keeping up with technology trends	6	6	7	1
14	Technology infrastructure	6,5	6	7	1
15	Easy access to data sources	6	6	7	1
16	Task - technology - people balance	6,5	6	7	1
17	Allocation of resources	6	6	7	1
18	Team management	6	6	7	1
19	Project schedule	7	6	7	1
20	Defining measurement of success	6	6	7	1
21	Documentation	7	6	7	1
22	Defining business needs and objectives	7	6	7	1
23	Suitability	6	5,5	6,5	1
24	Positioning Big Data within the enterprise	6	6	7	1
25	Compatibility with business processes	7	6	7	1

* This statement was obtained from the literature by the researcher and included in the second round.

Before the Delphi study, the researcher was conducted a literature review on CSF s of IT projects to gain a greater understanding of what topics have been researched in previous projects. There were several CSFs, which are found critical in most of the studies; upper management support (Gomez and Heeks, 2016; McAfee and Brynjolfsson, 2012; Saltz and Shamshurin, 2016; Yeoh and Popovic, 2016; Gao et al., 2015; Nieder, 2016; Chen et al., 2016; Kamioka and Tapanainen, 2014; Cao and Duan, 2014; Wamba et al., 2016; Ji-Fan Ren, et al., 2016; Garmaki et al., 2016; Wang and Byrd, 2017; Kim and Park, 2016; Dutta and Bose, 2015; Koronois et al., 2014) was one of these CSFs. Instead of the 24 CSFs emerged in the first round, the researcher also included this well-known CSF from the literature and sent 25 statements to panelists in round two.

At the second round, seventeen experts rated these elements based on the importance on a scale of 0- 7 with zero (0) indicating “strongly disagree” and seven (7) indicating “strongly agree”. Both the median and interquartile range were calculated for each element listed. An interquartile range of 2 or less demonstrated that consensus was reached for this element, and the lower the interquartile range, the greater the consensus among the expert panelists. The percentage of experts' responses that fell within the interquartile range was also calculated. At the end of the second consensus was reached for all items with IQR ≤ 1. The ratings and related statistics are shown in the table 8.

CHAPTER 5: BIG DATA PROJECT SUCCESS MODEL

For the quantitative portion, the method of collecting information was survey research which is conducted through CATI (computer-assisted telephone interview). The intent was to measure characteristics representing the population using statistical techniques (Kerlinger and Lee, 2000; Creswell and Plano Clark, 2013). Survey research is a convenient method of understanding aspects of behavior through statistical analysis from a sample of the population. The statistical analysis allows comparisons and strength of relationships between variables to the hypotheses being tested.

5.1. Scale Development

A survey instrument is a tool to predict using a sample to infer observations to a greater population (Neuman, 2003). It is also used to collect standardized scores by asking every person the same question in addition to collecting time- or context-specific data (Neuman, 2003).

The consensus within the literature is not apparent regarding the criteria to measure project success (Pinto and Prescott 1988). Modern theory on project performance consists of three dimensions: quality, time and cost (Guan and He, 2007). Measuring project success of performance is generally based on whether or not the project was completed on schedule, whether or not the project was completed within budget, and whether or not the delivered product benefits stakeholders (Cook, 2004; Gallegos et al., 2004; Kerzner, 2003; PMI, 2015). The perspectives in the literature were categorized as either project management behaviors or organizational behaviors. Therefore, one of the objectives of this study is to develop a reliable, valid, and generalizable scale that measures the CSFs of Big Data projects.

5.1.1. Methodology

The researcher followed well-accepted procedures for the conceptual development of factor identification (Hair et al., 2012) and the scale development process (Churchill, 1979; Crocker and Algina, 1986; DeVellis, 2016; Gerbing and Anderson, 1988; Netemeyer, Bearden, and Sharma, 2003; Nunnally and Bernstein, 1994) which are found from a review of the current literature. This process involves construct definition, item generation, and purification, content validity, reliability and validity assessments. The process involved an inductive approach by relying on qualitative analysis to generate scale items to measure the constructs.

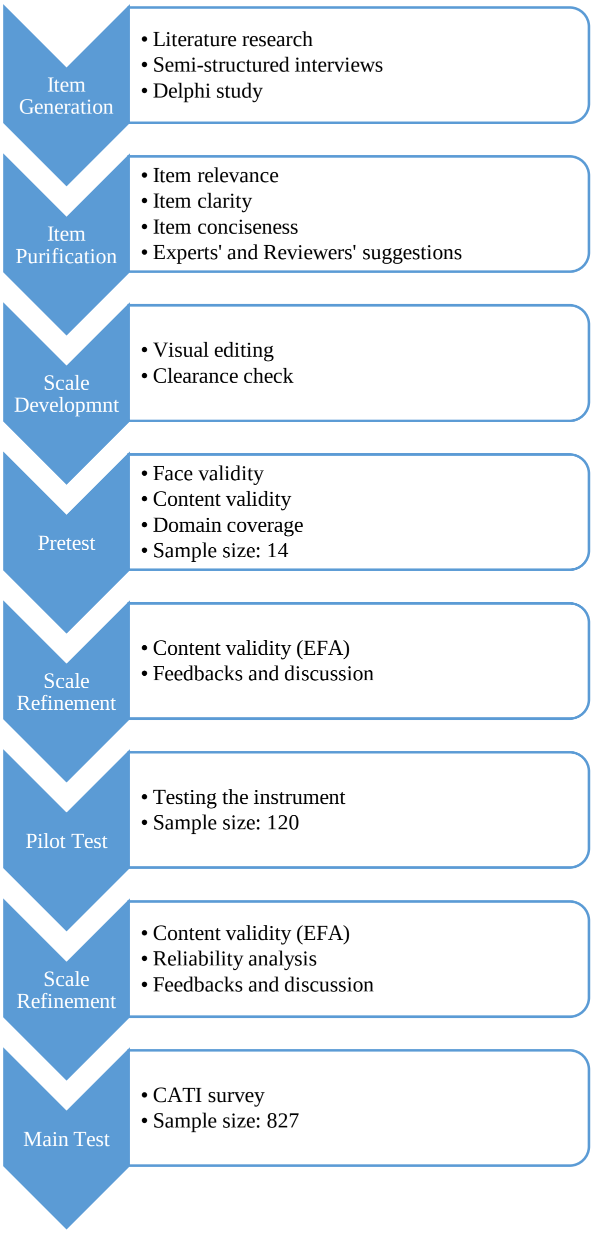

Figure 10: Scale Development Process

The figure illustrates the process followed for scale development. The items are extracted from the semi-structured interview and Delphi study results as described in the previous chapters. On the other hand, the literature frequently contained material addressing the topic of measuring several factors in accordance with project success and CSFs. These materials were not developed especially for Big Data projects, but as Big Data is also an IT project, the appropriate items found in the literature are matched with qualitative findings of this study. The developed scale measured project success and CSFs, based on an aggregate of 39 statements.

5.1.2. Construct Definition

This phase of scale development requires specificity in delineating the construct's domain and facets, and in establishing what the construct does or does not entail (Churchill, 1979; Zaichkowsky, 1985; Haynes, Nelson, and Blaine, 1999; Haynes, Richard, and Kubany, 1995; Nunnally and Bernstein, 1994). The construct domain may be specified via a literature review of related constructs and measures (Clark and Watson, 1995; Haynes et al., 1999). Importantly, the measure for the construct must possess content validity and be appropriate for reliably and accurately predicting behaviors.

As this is an exploratory research in nature, the constructs which will emerge at the end of EFA and after in SEM is not known/predicted at this phase of the research. Therefore, construct definition stage of the scale development flow is omitted.

5.1.3. Item Generation and Analysis

Item generation involves generating a representative pool of items for each dimension of the construct (Churchill, 1979). Often, open-ended responses are converted into items for different dimensions (Richins and Dawson, 1992; Shimp and Sharma, 1987). It is important to develop items that are clear, concise, and specific (Peterson et al., 1999; Podsakoff et al., 2003; Spector, 1992), and to purge items that are verbose, obscure, or confusing (Angleitner and Wiggins, 1985). The extant literature is examined to uncover additional scale items, which are then incorporated with the other items to comprise the initial set of items (Bearden, Hardesty, and Rose, 2001). In this stage of scale development, validity means value defined as important, interesting, or useful (McGrath and Brinberg, 1983).

5.1.3.1. Exploratory Qualitative Item Extraction

The goal of the qualitative portion was to discover CSFs that influence successful Big Data projects. Furthermore, this discovery would lead to a model being induced to allow us to test and establish statistical relationships between the variables. To accomplish the above goal, the extraction of variables from the Delphi study was completed in a total of seven steps, as described below, to induce a model for testing. In the first step, the researcher specifically extracted and categorized criteria of success. In the next six steps, all items that can impact successful Big Data projects were extracted, categorized and examined for patterns. Details around factors extracted are all listed in Chapter 4 as part of findings and analysis. The method followed for the extraction is as follows:

1. Raw data review to define "how does a successful Big Data project appear as".

This is what the result is at the end of a project. The research only focused on successful projects. In other words, to find the characteristics that impact successful Big Data projects, which are the research objective, we need to identify what success means for Big Data projects. This identifiable success will be what we measure in our model. This would be our dependent variable or outcome in the model. For instance, a new resulting product due to a Big Data initiative would be classified as a success item to measure for Big Data Implementations. The researcher identified these at the end of the semi-structured interview section.

2. Revision of codes to fit into categories.

In this round, the researcher reviewed all the identified codes, which is listed at the end of the second round of Delphi study and at the end of the semi-structured interview section, to see if they could be grouped into categories. Categories represent similar themes. For instance, in the Delphi study, the experts listed the importance of communication or the communication skills in their project. Through these separate examples, they represent that communication is a CSF for a successful Big Data project.

3. Revision of the categories to see patterns and themes for measurement.

Finally, in this third round, the researcher reviewed for patterns or major themes that emerged from raw data, codes, and categories.

Moreover, everything else that impacts the success item would be deemed an independent variable. An independent variable is a variable on which its variation or result is not dependent on another item. It stands alone and is not impacted by other variables. For instance, age is an independent variable which does not depend on anything for its result. Similarly, after the above two rounds of understanding what are the CSFs of a successful Big Data project means, we inspect the Delphi feedbacks for what will critically affect this success. In other words, we want to find what critically causes success in Big Data projects.

After having completed rounds one and two to identify success criteria and complete a list of CSFs, the next step was to find what such success actually depends on.

4. Raw data review to find keywords that call out characteristics that can impact the success of Big Data projects.

Here the researcher did a full review to find any characteristic that can impact successful Big Data projects. For instance, the experts listed about the focus on information, data and leadership abilities in terms of impact. All these were quotes, were extracted and identified as codes, just like in round 1. This first pass was to capture everything possible that would help answer the qualitative research question.

5. Review of the codes to fit into categories.

Major themes would act like broad categories where codes can fit in and can be grouped together to represent the initial set of codes. Similar in nature to round 2 but this time focusing on the codes developed from round 3, above. The only difference here is that context was important to consider and extract as well. For instance, when experts mention focussing on information and data it might be easier to group them together but keeping context in mind resulted in different categories being identified for each. One being, where the focus on information gathering was important versus ownership of data. As you can see both are different concepts and thus put in different categories.

6. Review for patterns or themes

Finally, in this last round, the researcher reviewed for patterns or major themes that emerged from raw data, codes, and categories. For instance, the expert listed about skills, tools, and platform from a technology standpoint. As such the major theme, here, could be classified as technical competence of the organization.

7. Conversion of codes into scalable items.

The researcher extracted a list of CSFs at the end of the previous 6 step process. Our final goal is to create a reliable scale to examine the relationship between the CSFs. The extracted codes are converted to survey items at this last step. For example, the code of “top management support” is converted as “Top management support has critical importance” in order to measure the level of agreement of the participants to each CSF item via Likert scale.

Table 9: Category Itemization

Category	Item
Financial efficiency	It is critical that the enterprise has the opportunity to invest in resources (technology, people, etc.) needed for the project.
Change management	Ability to manage changes in technology, people, task and structure is critical within the scope of the project.
Recruitment strategy	It is critical to recruit appropriate and right people who fit the project needs.
Top management support	Top management support is critical for the project.
Team leader skills	It is critical that the project team leader has managerial abilities.
Multidisciplinary team	It is critical for the project team to involve employees from each relevant department.
Team skills	It is critical that people in the project team have analytical thinking skills.
	It is critical that people in the project team have the necessary technical skills.
Education status of the team	It is critical that the project team is trained or educated on Big Data.
Communication ability	It is critical that people in the team communicate with each other in a healthy way.
	Healthy interaction between project team end users is critical.
Technology infrastructure	It is critical that the enterprise has a flexible IT infrastructure.
Defining technology strategy	Defining the strategy about which implementation and development tools will be used within the project is critical.
Data quality	It is critical that the data used in the project are qualified (complete, consistent, accurate, appropriate...).

Category	Item
Keep up with technology trends	The use of current analysis tools within the project is critical.
Technology infrastructure	The integration of old and new databases in the project is critical.
	It is critical that data management and auditing activities are performed hassle-free.
	It is critical to establish the IT infrastructure of the enterprise considering future needs.
Easy access to data sources	It is critical to access internal and external databases during the project.
Task - technology - people balance	Providing technology-task-people balance is critical in the project.
Allocation of resources	It is critical that the resources allocated to the project (human, technology, money, etc.) are properly distributed.
Team management	Proper job descriptions within the scope of the project are critical.
Project schedule	It is critical that the project schedule is clear.
	It is of critical importance that the project is progressively benefited in a short time (milestones are quickly accessible).
	It is critical that the project will be concluded and implemented in a short time.
Defining measurement of success	It is critical that the project has measurable or definable outputs.
Documentation	It is critical to document during the project to aware of know-how loss.
Defining business needs and objectives	It is critical that the business problem that the project aims to find a solution is well defined.

Category	Item
	Accurate identification of the purpose and scope of the project is critical.
	It is critical that the project is parallel to the business objectives.
Suitability	It is critical that the project scope fits the needs of the enterprise.
Positioning Big Data within the enterprise	It is critical to determine the strategic position of Big Data in achieving the objectives of the business.
Compatibility with business processes	It is critical that Big Data plays an important role in business and decision-making processes.
Reaching project goals	The project objectives were successfully reached.
Satisfaction of end users	The project end users were satisfied.
Reaching project budget targets	The project has achieved the budget target.
Reaching quality goals	The project has reached its quality target.
Reaching project schedule	The project was completed in accordance with the project schedule.
Perception of success	I believe the project was successful.

Each item was written to reflect the CSFs emerged through the qualitative study. Redundancy in items has both its pros and cons; while the final instrument should aim to lessen redundancy, it can make sense in the initial item pool. Of course, irrelevant redundancies should be avoided, i.e., those pertaining to incidental vocabulary and grammar (DeVellis, 2012).

One thing to consider with Likert scales is that overly mild statements might elicit too much agreement. The researcher should imagine how individuals from the target population with different strengths of the attribute or attitude in question are likely to respond. A measure among respondents cannot co-vary if it does not vary, therefore, items should be written so that variation among respondents is a reasonable expectation. Similarly, the number of response choices should be sufficient to allow for variation (i.e., six or seven)— but not so numerous that differences between response choices become meaningless (DeVellis, 2012). Therefore, the Big Data Success Scale will consist of statements in which participants will respond to their level of agreement on a scale of 1 to 7.

A number of questions in the semi-structured interviews and Delphi study resulted in concepts that were then categorized. These concepts and categories were a useful starting point in organizing the writing of scale items.

Table 10: Items Derived from the Qualitative Study

No	Items
1	It is critical that the enterprise has the opportunity to invest in resources (technology, people, etc.) needed for the project.
2	Ability to manage changes in technology, people, task and structure is critical within the scope of the project.
3	It is critical to recruit appropriate and right people who fit the project needs.
4	Top management support is critical for the project.
5	It is critical that the project team leader has managerial abilities.
6	It is critical for the project team to involve employees from each relevant department.
7	It is critical that people in the project team have analytical thinking skills.
8	It is critical that people in the project team have the necessary technical skills.
9	It is critical that the project team is trained or educated on Big Data.
10	It is critical that people in the team communicate with each other in a healthy way.
11	Healthy interaction between project team end users is critical.
12	It is critical that the enterprise has a flexible IT infrastructure.
13	Defining the strategy about which implementation and development tools will be used within the project is critical.
14	It is critical that the data used in the project are qualified (complete, consistent, accurate, appropriate...).
15	The use of current analysis tools within the project is critical.
16	The integration of old and new databases in the project is critical.
17	It is critical that data management and auditing activities are performed hassle-free.

No	Items
18	It is critical to establish the IT infrastructure of the enterprise considering future needs.
19	It is critical to access internal and external databases during the project.
20	Providing technology-task-people balance is critical in the project.
21	It is critical that the resources allocated to the project (human, technology, money, etc.) are properly distributed.
22	Proper job descriptions within the scope of the project are critical.
23	It is critical that the project schedule is clear.
24	It is of critical importance that the project is progressively benefited in a short time (milestones are quickly accessible).
25	It is critical that the project will be concluded and implemented in a short time.
26	It is critical that the project has measurable or definable outputs.
27	It is critical to document during the project to aware of know-how loss.
28	It is critical that the business problem that the project aims to find a solution is well defined.
29	Accurate identification of the purpose and scope of the project is critical.
30	It is critical that the project is parallel to the business objectives.
31	It is critical that the project scope fits the needs of the enterprise.
32	It is critical to determine the strategic position of Big Data in achieving the objectives of the business.
33	It is critical that Big Data plays an important role in business and decision-making processes.
34	The project objectives were successfully reached.
35	The project end users were satisfied.
36	The project has achieved the budget target.
37	The project has reached its quality target.
38	The project was completed in accordance with the project schedule.
39	I believe the project was successful.

5.1.3.2. Literature Review Item Generation

Other potential items can be generated from the literature on related theories or prior studies conducted on this issue. Because the CSFs of Big Data projects are not investigated before in terms of a quantitative research via survey, there are no items generated from the relevant literature.

There are several CSF research conducted on IT related topics. The items used in this research are examined as reference items to strengthen the items emerged within the qualitative part of this research.

Table 11: Item – Reference Mapping

Category	Item	Reference
Financial efficiency	1	Wamba et al., 2017; Chen et al., 2015; Kim and Park, 2016; Dutta and Bose, 2015; Janssen et al., 2017; Koronois et al., 2014; Audzeyeva and Hudson, 2016; Cleland and King, 1983; Kamal, 2006
Change management	2	McAfee and Brynjolfssn, 2012; Manyika et al., 2011; Sadovskyi et al., 2014; Kamioka and Tapanainen, 2014; Cao and Duan, 2014; Baker et al., 1983
Recruitment strategy	3	Cao and Duan, 2014; Wamba et al., 2016; Ji-Fan Ren, et al., 2016; Garmaki et al., 2016; Pinto and Slevin, 1987
Top management support	4	Halaweh and Massry, 2015; Popovic et al., 2016; Audzeyeva and Hudson, 2016; Arnott, 2008; Dawson and Van Belle, 2013; Green, Rutherford and Turner, 2009; Grubljesic and Jaklic, 2015; Hasan, Lotfollah and Negar, 2012; Hawking and Sellitto, 2010; Khojasteh, Ansari and Abadi, 2013; Salmasi, Talebpour and Homayounvala, 2016; Kohnke, Wolf and Mueller, 2011; Lautenbach, Johnston and Adeniran-Ogundipe, 2017; Nasab, DSelamat and Masrom, 2015; Olbrich, Poeppelbuss and Niehaves, 2011; Olszak and Ziemba, 2012; Pham et al., 2016; Popovic, Turk and Jaklic,

Category	Item	Reference
		2010; Puklavec, Oliviera and Popovic, 2014; Sparks and McCann, 2015; Yeoh and Koronios, 2010; Martin, 1976; Cleland and King, 1975; Bailey and Pearson, 1983; Pinto and Slevin, 1988, 1989; Standish Group, 1994; Jiang et al., 1996; Li, 1997; Murray, 2001; Baccarini and Collins, 2003; Dong et al., 2004; Wong and Tein, 2004; Kamal, 2006
Team leader skills	5	Gupta and George, 2016; Cao and Duan, 2014; Wamba et al., 2016; Ji-Fan Ren, et al., 2016; Garmaki et al., 2016; Wang and Byrd, 2017; Kim and Park, 2016; Dutta and Bose, 2015; Koronois et al., 2014; Verma, 1995; Turner and Müller, 2004, 2005; Freedman and Katz, 2007
Multidisciplinary team	6	Gomez and Heeks, 2016; McAfee and Brynjolfsson, 2012; Saltz and Shamshurin, 2016; Yeoh and Popovic, 2016; Gao et al., 2015; Nieder, 2016; Standish Group, 1994; Wixom, 2001; Baccarini and Collins, 2003; Dong et al., 2004; Wong and Tein, 2003; Kamal, 2006; Pinto and Slevin, 1988; 1989; Jiang et al., 1996
Team skills	7	Gupta and George, 2016; Saltz and Shamshurin, 2016; Sadovskyi et al., 2014; Wamba et al., 2017; Kamioka and Tapanainen, 2014; Garmaki et al., 2016; Wang and Byrd, 2017; Popovic et al., 2018; Dutta and Bose, 2015; Pospiech and Felden, 2016; Koronios et al., 2014
	8	Gupta and George, 2016; McAfee and Brynjolfsson, 2012; Yeoh and Popovic, 2016; Gao et al., 2015; Nieder, 2016; Chen et al., 2016; Pinto and Slevin, 1988, 1989; Jiang et al., 1996; Wong and Tein, 2004
Education status of the team	9	Gupta and George, 2016; Chen et al., 2015; Baker et al., 1997; Jiang et al., 1996; Baccarini and Collins, 2003; Dong et al., 2004

Category	Item	Reference
Communication ability	10	Gupta and George, 2016; Chen et al., 2015; Lock, 1984; Cleland and King, 1983; Pinto and Slevin, 1988, 1989; Verma, 1995; Jiang et al., 1996; Baccarini and Collins, 2003; Sofian, 2003; Dong et al., 2004; Wong and Tein, 2004; Reel, 1999; Wixom, 2001; Baccarini and Collins, 2003
	11	Bailey and Pearson, 1983; Li, 1997; Gupta and George, 2016; McAfee and Brynjolfsson, 2012; Yeoh and Popovic, 2016; Lock, 1984; Cleland and King, 1983; Pinto and Slevin, 1988, 1989; Verma, 1995; Jiang et al., 1996; Baccarini and Collins, 2003; Dong et al., 2004; Wong and Tein, 2004
Technology infrastructure	12	Halaweh and Massry, 2015; Manyika et al., 2011; Janssen et al., 2017; Wamba et al., 2017; Wang and Byrd, 2017
Defining technology strategy	13	Wang and Byrd, 2017; Saltz and Shamshurin, 2016; Sadovskyi et al., 2014; Kamioka and Tapanainen, 2014; Garmaki et al., 2016; Wang and Byrd, 2017; Popovic et al., 2016; Dutta and Bose, 2015; Pospiech and Felden, 2016; Koronois et al., 2014; Kim and Park, 2016
Data quality	14	Wang and Byrd, 2017; Qwan et al., 2014; Akter et al., 2016; Garmaki et al., 2016; Dutta and Bose, 2015; Abbas and Aggarwal, 2010; Halaweh and Massry, 2015; Kim and Park, 2016
Keep up with technology trends	15	Janssen et al., 2017; Gupta and George, 2016; Dutta and Bose, 2015; Cao and Duan, 2014; Kim and Park, 2016
Technology infrastructure	16	Wang and Hajli, 2017; Abbas and Aggarwal, 2010; Koronios et al., 2014; Kim and Park, 2016; Wong and Tein, 2004

Category	Item	Reference
	17	Wang et al., 2018; Wang and Hajli, 2017; Gomez and Heeks, 2016; Kamioka and Tapanainen, 2014; Ji-Fan Ren, et al., 2016; Wang and Byrd, 2017; Kim and Park, 2016;
	18	Cao and Duan, 2014; Wang and Byrd, 2017; Wang et al., 2016; Gupta and George, 2016; Kim and Park, 2016; Pospiech and Felden, 2016; Koronios et al., 2014
Easy access to data sources	19	Cao and Duan, 2014; Gomez and Heeks, 2016; Kamioka and Tapanainen, 2014; Ji-Fan Ren, et al., 2016; Wang and Byrd, 2017; Wang et al., 2016; Kim and Park, 2016; Pospiech and Felden, 2016; Koronios et al., 2014
Task - technology - people balance	20	Arnott, 2008; Fourati-Jamoussi, 2016; Khojasteh et al., 2013; Dutta and Bose, 2015; Olszak and Ziemba, 2012; Ravasan and Savoji, 2014
Allocation of resources	21	Dutta and Bose, 2015; McAfee and Brynjolfssn, 2012; LaValle et al., 2011; Manyika et al., 2011
Team management	22	Popovic et al., 2016; Dutta and Bose, 2015
Project schedule	23	Ji-fan et al., 2016; McAfee and Brynjolfssn, 2012; LaValle et al., 2011; Manyika et al., 2011
	24	McAfee and Brynjolfssn, 2012; LaValle et al., 2011; Manyika et al., 2011
	25	McAfee and Brynjolfssn, 2012; LaValle et al., 2011; Manyika et al., 2011
Defining measurement of success	26	Ji-fan et al., 2016; Kohnke, Wolf and Mueller, 2011; Lautenbach, Johnston and Adeniran-Ogundipe, 2017; Nasab, DSelamat and Masrom, 2015; Olbrich, Poeppelbuss and Niehaves, 2011
Documentation	27	Ji-fan et al., 2016; LaValle et al., 2011; Manyika et al., 2011

Category	Item	Reference
Defining business needs and objectives	28	Popovic et al., 2016; Dutta and Bose, 2015; Ji-fan et al., 2016
	29	LaValle et al., 2011; Manyika et al., 2011
	30	Popovic et al., 2016; Wang and Byrd, 2017; Wang et al., 2016; Janssen et al., 2017; Halaweh and Massry, 2015
Suitability	31	Wang and Byrd, 2017; Wang et al., 2016; Janssen et al., 2017; Halaweh and Massry, 2015
Positioning Big Data within the enterprise	32	Adamala and Cidrin, 2011; Hackney et al., 2015; Salmasi, Talebpour and Homayounvala, 2016; Kulkarni, Robles-Flores, 2013; Ravasan and Savoji, 2014; Sparks and McCann, 2015; Yogev, Even and Fink, 2013
Compatibility with business processes	33	Janssen et al., 2017; Popovic et al., 2016; Hackney et al., 2015; Kulkarni and Robles-Flores, 2013; Mudzana and Maharaj, 2015; 2017; Nemec, 2011
Reaching project goals	34	Wang and Hajli, 2017; Popovic et al., 2016; Pospiech and Felden, 2016
Satisfaction of end users	35	Kwon et al., 2014; Kohnke, Wolf and Mueller, 2011; Gaardboe, Nyvang and Sandalgaard, 2017; Gonzales, Wareham and Serida, 2015; Hackney et al., 2015; Kulkarni and Robles-Flores, 2013; Mudzana and Maharaj, 2015; 2017; Nemec, 2011; Olszak and Ziemba, 2012; Tona ,Carlsson and Eom, 2012; Visinescu, Jones and Sidorova, 2017; Wieder, Ossimitz and Chamoni, 2012; Standish Group, 1994; Reel, 1999; Wong and Tein, 2004
Reaching project budget targets	36	Kohnke, Wolf and Mueller, 2011; Adamala and Cidrin, 2011; Hackney et al., 2015

Category	Item	Reference
Reaching quality goals	37	Kwon et al., 2014; Wang and Hajli, 2017; Popovic et al., 2016; Pospiech and Felden, 2016
Reaching project schedule	38	Kwon et al., 2014; Wang and Hajli, 2017; Popovic et al., 2016; Pospiech and Felden, 2016
Perception of success	39	Kwon et al., 2014; Wang and Hajli, 2017; Popovic et al., 2016; Pospiech and Felden, 2016; Nemec, 2011; Kohnke, Wolf and Mueller, 2011

5.1.3.3. Pretesting

A preliminary pool of items is developed based on qualitative research. Pre-testing was conducted to assess whether any items were ambiguous or confusing (Dillman, Smyth, and Christian, 2014; Spector, 1992), which occurred via expert and academician reviews. Good items are clear, unambiguous, contain a single idea, and are not overly long or wordy; the reading difficulty level should be taken into account (Devellis, 2012; Spector, 1992). Pretests of these items are conducted and correlations, descriptive statistics, and item analyses are examined to identify low communalities, any ambiguity in the language used, and potential complexities in wording that might represent more than one underlying concept in a single item.

Agreement response anchors are versatile and popular, with 5 to 9 choices optimal (Spector, 1992) and were used here. The response choices for each item were: 1=Strongly disagree, 2=Somewhat disagree, 3=Disagree, 4=Neutral, 5=Agree, 6=Somewhat agree, 7=Strongly agree.

Based on these early statistical explorations and conceptual item analyses, changes are made to the initial set of items to enhance the starting pool and more accurately sample the domain of interest.

The pool of items is sequentially presented separately to each of the nine Big Data experts, who are asked to provide feedback on the items. The aim is to obtain a preliminary check of face validity, content validity, and thoroughness of domain coverage. Items are modified, created, and/or rejected based on the comments of these informants. Following this exploratory process, the remaining items are used to collect subsequent quantitative data to explore the structure of the item pool and to undertake item purification using accepted scale development practices.

The key informant technique is then applied to the improved item pool as a continuation of this exploratory research (Parasuraman, Grewal, and Krishnan, 2006). Before the survey was distributed to the respondents, face validation was conducted with five academicians and nine experts. There were 2 main goals for pre-testing the survey: a clear understanding of the survey content/questions and overall survey functionality (ease of, total time, format and flow of survey). Also, this was to ensure that the questions asked were relevant to this study especially with regards to each variable and topic of interest as well as applicability to the Big Data projects in the given sector. Additionally, it was to ensure that the items used were good measurements that would enable the respondents to relate to their situation and also to ensure that the respondents have no difficulties with the technical or organizational terms used in the survey.

The five academicians who reviewed the questions have knowledge in information systems and have experience in developing scales and conducting semi-structured interviews. The nine experts have been working for more than four years in the Big Data industry and these experts were different from the experts who are recruited for semi-structured interviews and Delphi study.

The questionnaire was given to them and they were asked to fill it up. Comments and feedbacks were requested. The respondents were also asked if any of the terms used in the survey needed clarification. One of them suggested changing from having the respondents' write down the answer (ratio scale) to selecting the answer based on choices provided (ordinal scale) for the question identifying the respondent's status, which requests information on revenue and number of employees as she thought that some respondents may not be willing to reveal the information as it may be considered as private and confidential to the firm. Apart from that, 2 questions were corrected due to ambiguity as suggested by the respondents during the pre-testing. The adjustment was carried out accordingly. No adjustment was made thereafter.

5.1.3.4. Pilot Study and Exploratory Examination

Balian (1994) recommends a pilot study on the basis that it provides the researcher a full review of the questionnaires, respondents and actual test administration. Thus, a pilot study was conducted via online survey as it provided the feel of the data and the opportunity to objectively measure the validity and reliability of the instrument.

Reliability refers to the extent of consistency of a measurement of a concept over time (Sekaran and Bougie 2010). It means the measures of a particular concept are able to produce stable and consistent results on repeated trials so that the results do not fluctuate. The purpose of conducting reliability test is to determine the extent of the reliability of each item specified in the respective constructs based on the scores indicated by the respondents (Sekaran and Bougie 2010) and to determine whether the items that measure a particular construct can be grouped together (Pallant, 2007). Reliability analysis was ascertained through examining the Cronbach's Alpha to decide which items were to be dropped and which were to be kept. Prior to using Structural Equation modeling to analyze the structural model, a preliminary reliability analysis was performed on the pilot data of 120 responses, which are gathered from Big Data experts.

Table 12: Reliability Analysis for Pilot Study

	Cronbach's Alpha
Governance	0,752
Project Definition	0,796
Project Management	0,823
Success	0,875
Team	0,805
Technology	0,873

The responses used for the pilot testing are sufficient as Cooper and Schindler (2003) suggested 25 to 100 respondents for pilot testing. All the reflective constructs were subject to reliability test using Cronbach's alpha. The purpose is to check the internal consistency of each latent construct so that it is clearly defined and captured. Items will be considered for removal if such act results in a significant increase in the Cronbach's alpha value. The results demonstrated that all the reflective constructs have sufficient internal consistency, with all above the value of 0.7, which is the cut off suggested by Nunnally (1978). The results of the reliability analysis of the pilot study data are provided in Table 12.

During pre-testing and pilot study (June 3rd to June 15th, 2018), it was observed that the average time to complete the survey was less than 20 min. The survey initially had a total of 50 questions of which 6 were demographics and 2 were conditional questions designed to ensure reliability and accuracy of the data gathered to measure successful Big Data projects.

Upon feedback, 10 survey questions had the wording changed for better focus on single constructs. Other edits, including, randomization of the order of questions, required answering of certain questions and the addition of the “other” choice was made. The survey completion time has decreased to less than 15 minutes with these minor changes.

5.1.4. Item Purification

Item purification is undertaken to ensure that the developing scale is measuring what it is intended to measure and to further refine the item pool (Shimp and Sharma, 1987; Bearden, Hardesty, and Rose, 2001). Factor analysis is utilized to reduce data and refine a developing scale (Ford, MacCallum, and Tait, 1986). Items are eliminated based on several criteria, including factor loadings, the correlation (or regression weight) of a variable with a factor, inter-item correlations, and item-to-total subscale correlations to determine if the items have statistically high correlations with their intended dimension. The outcome of item purification did not reduce the set of items.

Exploratory factor analysis (EFA) allows a researcher to discover the nature of the constructs influencing a set of responses by statistically determining the number of common factors (sometimes called dimensions) influencing a scale or set of measures. Factor axes are rotated in order to obtain simple and interpretable factors (Yaremko, 1986). The objective of EFA is to maximize the percent of variance explained by the model that it lays out. The researcher does not need to have a model in mind at the onset of EFA; factors are derived from the data and then interpreted by the researcher. Exploratory factor analysis also helps determine the structure of the items.

Any factor loading greater than 0.5 is assumed to possess practical significance (Hair et al., 2010) any item not demonstrating practical significance is eliminated. Item elimination also occurs if the item demonstrates significant cross-loadings (above 0.4) between two or more factors and non-significant loadings (less than 0.5) on any one factor. The decision of the appropriate number of factors is based on a combination of conceptual foundation and empirical evidence (Hair et al., 2010).

The results of EFA allow for interpretable sets of items that group together, have some practical relevance and are empirically supported (Hair et al., 2010). The constructs are then named and investigated further. Empirically, a solution with more than 50% of variance explained and communalities of 0.4 or higher is deemed appropriate. This is an iterative process, in that the loadings and communalities change as items are eliminated. The final solution results from multiple iterations of item analysis. Once a final solution is reached, in the light of the literature review, the items are referenced from similar studies investigating relevant projects or tools.

5.2. Research Timeline

The timeline for the quantitative part was from June 2018 to August 2018. The pilot testing was conducted from June 3rd to June 15th, 2018. The CATI survey was conducted from June 25, 2018, to July 15, 2018.

5.3. Process and Compilation

A private firm was hired by the researcher to conduct the survey. The survey administered through a CATI system. It included an IT workers sample. The use of survey method with CATI technique enabled the researcher to gather information nationally. The survey comprising of a total of 50 questions scaled with multiple choice and the Likert scale was conducted through telephone. Confidentiality was maintained to not tie responses that would identify individuals there were. Of the 50 questions; 9 of were demographics related, 2 were targeted conditional questions used as a check to ensure the individual can contribute to the survey and the remaining 39 were questions surrounding variables captured from our previous item extraction studies. The variety of perspectives shared in the semi-structured interviews made clear that to gather meaningful data and use it for research in successful Big Data implementations the researcher had to make sure to ask two conditional questions before a responder can take further part in the survey. The reasoning behind the 2 targeted conditional questions was to maximize data collection from organizations/individuals that knew what Big Data was and had the experience with Big Data projects. Only implementing part one of the above seemed incomplete given that the research's objective was to study successful Big Data projects. It was important, however, to understand Big Data before the act of implementation. While the researcher understood that the understanding for Big Data can be subjective the implementation of such was less so and hence the addition of the second question to ensure the quality of responses captured this aspect for this research. Also we wanted the participants to rate the survey items in accordance with their “last” big data project. The main purpose of this instruction was because the last project is probably the most clearly remembered project so the instability would be eliminated.

Given the above condition, the CATI survey was divided into 2 parts asking the two conditional questions. If the response to one of the conditional questions was no, the survey stopped there and the surveyor thanked the participant for their time.

5.4. Sampling

As in the qualitative chapter, purposeful sampling is chosen for the quantitative part of this research. Patton described that the logic and power of purposeful sampling is due to selecting information-rich individuals as attendants (Patton, 1990). This allows the research team to learn the most about the central importance to the purpose of the evaluation, thus the term “purposeful sampling” (Coyne, 1997). The first step in conducting an evaluation using the purposeful sampling methods is to identify the characteristics of the sample and document the rationale for studying them. This will help the researcher describe the context of the program evaluation.

The sampling frame is the population relevant to the study (Daniel, 2012). Because this study has a “niche topic related” research question, the questionnaire should be replied by experts or workers of the related field. This study’s sampling frame is IT workers directory and the study is conducted only on individuals who had Big Data project experience before and so has the ability to answer the questions by thinking about their last Big Data project experience.

5.5. Sample Size Determination

The sampling method and the sample size used in a quantitative study, which seeks to make a statistical based generalization from the study results to a wider population, is a significant piece of the study and must be appropriate (Daniel, 2012). It is the representativeness of the sample that allows a researcher to generalize the research results to a larger population (external validity). There is a lot of debate with regards to sample size, and even more so when Structural Equation Modelling (SEM) is employed. Sekaran (2003) suggested that a sample size of more than 30 and less than 500 shall be appropriate for most studies as a general rule of thumb. Bartlett, Kotrlik, and Higgins (2001) and Hair et al. (2010) on the other hand, suggested a sample size of more than 100, while Anderson and Gerbing (1988) and Bagozzi and Yi (1989) suggested a minimum sample size of 100 to 150, Some suggested any number above 200 as sufficient (Hoe, 2008) while a sample size of below 50 is not recommended (Hair et al., 2010). Tabachnick and Fidell (2007) considered a sample size bigger than 300 cases "comfortable" for inferential studies. Additionally, they offered the following guide for sample sizes in inferential studies: 50 cases as very poor sample size, 100 cases as poor sample size, 200 cases as fair sample size, 300 cases as good sample size, 500 cases as very good sample size, and 1000 cases as excellent sample size. Similarly, VanVoorhis and Morgan (2007) indicated that 300 cases were considered a good sample size when they recapitulated the rule of thumb for the determination of appropriate sample size for various types of statistical analysis aiming at generalizing their results.

Cohen (1988) suggested that for structured research planning, it is important to determine the optimal sample size to obtain a specified power for a chosen significance criterion and effect size. The four important parameters of statistical inference for statistical power analysis are: (a) the statistical power, (b) the significance criterion, (c) the sample size, and (d) the effect size. Cohen (1988) emphasized that determining the sample size as a function of the effect size, the significance criterion, and the statistical power must be at the core of any rational basis for deciding on the sample size to be used in an investigation. In inferential research planning, researchers are called to determine the sample size necessary to attain a desired statistical power for a specified significance criterion, also called Type I error (α), and a hypothesized effect size (ES). The statistical power of the significancy test is the long-term probability of rejecting the null hypothesis. It is a function of the research sample size, the significance criterion, and the population effect size. Cohen (1988) asserted that for any statistical inference model, the relationships among these four variables (sample size, statistical power, significance criterion, and effect size) are such that any one of them is a function of the three others. A statistical test power of a null hypothesis is the probability that it will result in the rejection of the null hypothesis. Therefore, it is the probability that it will lead to the conclusion that the phenomenon under examination exists.

Cohen's (1988) statistical power analysis is design to determine an adequate sample size that optimizes study results' precision. This study assumed that the effect of the independent variables on the dependent variable could be positive, null, or negative. A one-tail (directional test) by definition rejects the null hypothesis in the opposite direction of that predicted because it has no power to predict such an effect. A two-tail statistical test was used to test the five null hypotheses defined in this study. The rejection of the null hypotheses is a token that the phenomenon to be demonstrated is in fact present. The reliability or precision of a sample value, as measured by the sample's standard error, is the closeness with which it can be expected to approximate the relevant population value (Cohen, 1988). The larger the sample size, the smaller the error, and the greater the reliability or precision of the results. In addition, the greater the precision of the sample size, the greater the probability of detecting the phenomenon under examination will be.

Apart from those mentioned above, there is also the 10 times rule of thumb by Hair et al. (2013). Research model of this study consists of a total of six latent variables, with 37 arrowheads pointing to the latent variables.

Thus, this method required a minimum of 370 respondents (37 arrowheads pointing at latent variables in PLS path model x 10) to conduct this study. It is important to note that all these are only a rough estimation of the minimum sample size.

In conclusion, an increase in sample size increases the statistical power, which is the probability of detecting the phenomenon of interest. The optimal sample size for this study, based on six predictors and the parameters herein defined is 912 participants.

A total number of 912 responses are recorded via CATI from target respondents using the sampling design and sampling method discussed in previous sections. Before conducting any statistical analysis, a rigorous filtering process was conducted to ensure the data set is complete and usable. Recorded data is checked for outliers and missing values. Also, some forms were rejected because of straight lining, which refers to the answers given were all same response for a high percentage of the question (Hair et al., 2014). For instance, some questionnaires were found to have the same answers for all the questions such as all middle point (neutral), maximum point or minimum point. 85 forms were rejected, leaving 827 usable responses which are sufficient for structural equation modeling. The number of responses met the requirements according to both suggestions from the literature. A more accurate estimation should be done using statistical power analysis with the largest number of predictors (Roldán and Franco, 2012).

5.5.1. Statistical Significance Criterion (α)

The significance criterion denoted by α represents the probability of mistakenly rejecting the null hypothesis, hence committing Type I error. A conservative α could The significance criterion denoted by α represents the probability of mistakenly rejecting the null hypothesis, hence committing Type I error.

5.5.2. Effect Size

Without intending any implication of causality, the effect size is the degree to which the phenomenon under examination is present in the population. For multiple regression analysis, the effect size index (F2) for small, medium, and large effect sizes are .02, .15, and .35, respectively. The null hypothesis implies that the effect size is equal to zero. Thus, the larger the effect size value, the greater the degree to which the phenomenon under examination is present. Similarly, the larger the posited effect size, the greater the power of the test, other parameters (significance criterion and sample size) being equal. In addition, the greater the effect size posited, other parameters (significance criterion and desired statistical power) being equal, the smaller the sample size necessary to detect the phenomenon under examination. On the contrary, the smaller the effect size, the more difficult it would be to detect the degree of deviation of the null hypotheses in actual units of response. Cohen (1992) suggested that a medium effect size is desirable as it would be able to approximate the average size of observed effects in various fields. This study will set the effect size (F2) at a level of 10%.

5.5.3. Statistical Power

Statistical power is defined as the probability that the statistical significance test will lead to the rejection of the null hypothesis for a specified value of an alternative hypothesis (Cohen, 1988). Power analysis offers researchers the ability to reject the null hypothesis in favor of the alternative when there is sufficient evidence from the collected sample that the value of a population parameter of interest is different from the hypothesized value.

The statistical power is expressed as (1-β), where β is the probability of wrongly accepting the null hypothesis. Such acceptance of the null hypothesis is known as committing Type II error. The value of the statistical power can range from zero to one. High (2000) argued that when an investigation's statistical power is low, the risk of committing Type II error is higher as there is little chance of detecting a significant effect if it is present. However, when the power is set too high, a small variation in the effect is detectable and the results are significant, but the size of the effect is not practical. Cohen (1992) noted that larger statistical power resulted in high sample sizes that might exceed researchers' available resources for data collection. Statistical power should be maximized for a research design to be sensitive to the existence of the phenomenon under consideration because it represents the probability of making the correct decision about the existence of the phenomenon (Hedges and Rhoads, 2010; Lipsey, 1990). This research used a high statistical power of .99 (β = .01), as the data collection focuses on a niche topic and gathers data from information-rich individuals. Setting β at the level of .01 means that the chance of wrongfully concluding that there is no effect (acceptance of H_0) is 1%.

5.6. Exploratory Data Analysis (EDA)

Prior to inferential testing, Tabachnick and Fidell (2007) reasoned that all multivariate assumptions (normality, linearity, homoscedasticity, and independence of errors) should be verified through an Exploratory Data Analysis (EDA). Additionally, to ensure that the results are not bias, multicollinearity, and extreme values (outliers) should be verified and corrected (Mertler and Vannatta, 2005; Tabachnick and Fidell, 2007) before inferential data analysis. This research utilized SPSS version 21 to conduct an EDA and correct any multicollinearity and outlier issues prior to the inferential study.

In order to have an overview of the general characteristics of respondents, respondents' demographic profile is tabulated into table form. In advance of performing any statistical analysis, it is important to check for blank, inconsistent, illogical responses in the dataset. Outliers are a form of illogical response because it is an observation that is substantially different from the other observations. Furthermore, researchers may have the tendency to key in the codes wrong. For example, researchers might accidentally type in 8 or 9 for a construct measured in 7- point scale. Such a mistake is illogical because the maximum code entered should be only 7. To prevent such issues from happening, the data was scrutinized through frequency check (e.g, frequencies, percentages, means, medians and range), so that a complete and usable data can be obtained (Sekaran and Bougie, 2010).

After collecting a total of 912 responses, 85 were dropped due to unfulfilled requirements such as missing values, outliers, and unreliable data, hence resulting in a final sample size of 827. Data cleaning and data preparation were also performed in accordance with the guidelines by Pallant (2007). This included checks on the scores within the range. The scores beyond the range might be due to missing values; thus, double checking was carried out and any missing values were identified.

Apart from that, preliminary steps were also taken before hypothesis testing. Data collected was checked for extreme values or outliers, normality, multicollinearity, reliability, and consistency.

5.6.1. Missing Values

The CATI survey was programmed to end with no missing values. Accordingly, the data consists of zero missing values as planned and expected.

Data is examined to identify unreliable responses, where ALL the values for questions using Likert-scale measurements were responded as "3" or "NEUTRAL". This represents the unwillingness and disinterest of the respondents to participate seriously in the survey. The deletion of these cases was to ensure that their responses did not affect the findings.

5.6.2. Outliers

The presence of extreme values (also known as outliers) may affect the arithmetic mean (Kline, 2005) as outliers can distort the interpretation of mean. According to Tabachnick and Fidell (2007), the presence of outliers may be due to a few reasons, such as a human error in data entry, misspecification of missing value code, or perhaps the sample set may not be part of the population. The detection and elimination of outliers are very important as it can improve multivariate normality (Kline, 2005).

The continuous variables were all designed in standardized Likert scales between 1 and 5 and no reverse coded questions were required since the questionnaire was arranged to follow the outcomes from previous literature. Thus, the data did not consist of univariate outliers. Using simple "Conditional Formatting" Function in Excel, outliers which were due to human error during data entering were detected. The errors were corrected and no univariate outliers were removed from the data set.

Apart from that, an examination of the z-scores revealed three respondents that can be classified as univariate outliers with z-scores in excess of ± 3.29 (Hair et al., 2013). Two respondents were identified in the z-score of one of the items of Success and another was identified in the z-score of one of the items of Team. However, according to the guideline by Hair et al. (2013), the ± 3.29 cut off is considered an outlier for small sample sizes, i.e. sample sizes less than 80, Thus, deleting the three cases would seem to be weakly justified. Moreover, as the boxplot checks were run, two cases were identified as outliers and any extreme outlier is observed.

Pallant (2007) in his guideline advised for the removal of only the extreme outliers. Thus, the values were kept in the final dataset to be used in subsequent analyses.

In total, 85 cases were deleted. The final dataset then consisted of 827 samples.

5.6.3. Normality

Normally distributed data is described as having a symmetrical bell-shaped curve, where the greatest frequencies of scores are in the middle (Gravetter and Wallnau, 2016). Normality is tested only on the dependent variables. However, PLS-SEM does not require data to be normally distributed (Hair et al, 2013). This is one of the many advantages of using PLS as discussed previously. Even though PLS-SEM is a nonparametric statistical model, Hair et al. (2014) recommended that the data should not be too far from normal. Bootstrapping was used in the case of highly skewed data.

A Kolmogorov-Smirnov test and Shapiro-Wilks test for normality were used with skewness and kurtosis to check for normality using SPSS. According to the results and graph examinations, the distribution was significantly non-normal. Field (2009) explained that statistically significant results for the K-S test do not necessarily mean that the deviation from data normality is serious. Normality bias may be topically related, where the items were about CSFs and big data professionals agree with the importance of CSFs as it is expected. Accordingly, this causes left/negatively skewed distribution, where the responses accumulate on higher degrees of the Likert scale.

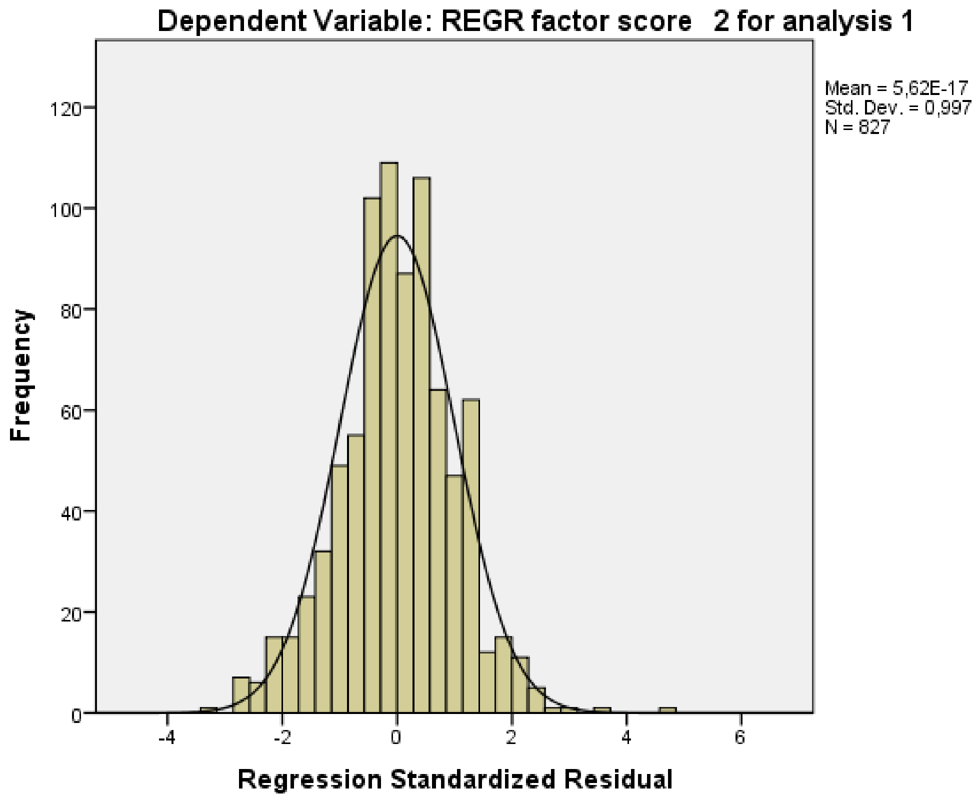

Figure 11: Histogram

The histogram in Figure 11 is plotted with the frequency of dependent variable (Success) vs. regression standardized residual. The shape of the histogram shows that the normality assumption is acceptable.

5.6.4. Multicollinearity (MC)

A high correlation between two or more predictor variables in a multiple regression model is a phenomenon known as multicollinearity. Since a high level of multicollinearity causes confusion and misleading results (Tabachnick and Fidell, 2012), it was assessed here using the Variance Inflation Factor (VIF) as recommended by (Tabachnick and Fidell, 2007).

Table 13: VIF Values

	Project Management	Success	Team	Technology
Governance	1.091		1.000	1.000
Project Definition	1.091	1.180		
Project Management		1.240		
Success				

	Project Management	Success	Team	Technology
Team		1.269		
Technology		1.397		

All variables in this study have VIF values below 3.3, the accepted criterion put forth by Diamantopoulos and Siguaw (2006). Hair et al. (2013) on the other hand, suggested a threshold of 5.0. Alternatively, inspecting the correlation matrix for independent variables is the simplest and most obvious method to find multicollinearity. The presence of high correlations of 0,90 or above suggests a substantial MC (Hair et al, 2010; Tabachnick, Fidell and Osterlind, 2001). The result shows that all the items have correlation values below 0,9; thus, it was concluded that there is no multicollinearity in the data.

5.7. Data Collection

The questionnaire was designed carefully in order to meet the optimal duration for CATI technique. An average interview lasted approximately 15 minutes and this duration is appropriate for an interviewer to stay concentrated and response to questions (Anie et al., 1996, Ketola and Klockars, 1999; Choi, 2004). After corresponding with crucial CATI criteria, the questionnaire was programmed into a computer-assisted telephone interviewing (CATI) system. The data was collected by a statistical research company. The CATI survey starts with capturing basic demographic data from participants. Aside from demographic and conditional questions, the surveyors asked each respondent to reply with 1=Strongly disagree, 2=Somewhat disagree, 3=Disagree, 4=Neutral, 5=Agree, 6=Somewhat agree, 7=Strongly agree. Participants indicated their response by saying the number of their level of agreement. Data collected through the CATI software was securely stored on servers. Data was further exported to SPSS and Smart PLS for analysis and kept confidentially.

5.8. Descriptive Statistics

The 827 responses used in this study is comprised of reliable and complete data without missing values, and outliers. The tables present the frequency and percentage break down of each demographic variable. The variables for the demographic profile or participant characteristics are the big data professionals' age, gender, title, educational level, working experience, the industry they worked, several educational details about IT department and business size.

Table 14: Sample Distribution by Industry

	Frequency	Percent
Education	27	3,3
Finance	109	13,2
Automotive	34	4,1
Energy	29	3,5
IT	231	27,9
Food	29	3,5
Wood	5	,6
Construction	39	4,7
Chemical-Plastic	50	6,0
Health	39	4,7
Electronic	40	4,8
Retail	97	11,7
Textile	50	6,0
Media-Communication	48	5,8
Total	827	100,0

The finance industry is one of the greatest investors of Big Data technologies. Industry segmentation of the sample highlights the big players of the Big Data market. IT companies may refer to outsource services about Big Data solutions. After the IT industry, finance with 13,2% and retail with 11,7% leads the segmentation. Table 14 presents the frequency and percentage of Big data professionals from different industries.

Table 15: Sample Distribution by Years of Big Data Experience

	Frequency	Percent
1-3 years	263	31,8
4-6 years	398	48,1
7-9 years	138	16,7
10 years or over	28	3,4
Total	827	100,0

The target population was Big Data professionals with at least one year of experience working on Big Data projects. Based on the four experience segments most of the respondents have 4-6 years of Big Data project experience. 10 years or over is the smallest group as expected. Big data has a short history in Turkey, the respondents in the most experienced group represent the pioneers of this field.

Table 16: Sample Distribution by Years of IT Experience

	Frequency	Percent
1-3 years	207	25,0
4-6 years	306	37,0
7-9 years	172	20,8
10 years or over	142	17,2
Total	827	100,0

We also asked the professionals their experience in the field of IT. Most of them have at least 4 years of experience. 207 participants reported that they have 1-3 years of experience, which represents 25% of the sample, 306 responses are recorded under 4-6 years, which refers to 37% of the population, 172 professionals indicated they have 7-9 years of experience on IT, which represents 20,8% of the sample and finally 142 participants indicated they were in the most experienced group with 10 years or over experience on the field and that refers to 17,2% of the sample.

Table 17: Sample Distribution by Gender

	Frequency	Percent
Female	215	26,0
Male	612	74,0
Total	827	100,0

For gender, there were 215 female and 612 male participants, which represented 26% and 74% of the sample, respectively. Table 17 presents this information in a summarized form.

Table 18: Sample Distribution by Age

	Frequency	Percent
18-25	123	14,9
26-35	387	46,8
36-45	215	26,0
46-55	102	12,3
Total	827	100,0

The survey included a demographic question for participant ages. This question asked participants to select the particular age group in which they belong. An analysis of the age groups shows that 14,9% identified themselves in the 18 – 25 age group, the largest age group was 26 to 35, which represented 46,8% of the responses, 26% of participants were in the 36 – 45 age group and finally 12,3% were in the 46 – 55 age group. Table 18 shows the frequency distribution for age in a summarized format.

Table 19: Sample Distribution by Education

	Frequency	Percent
Doctorate Degree	33	4,0
Master's Degree	70	8,5
Undergraduate	517	62,5
Associate Degree	207	25,0
Total	827	100,0

The survey included a demographic question about the participant's highest level of education. 33 participants indicated they have a Doctorate degree, representing 4% of the sample, 70 participants reported that they have a Master's degree, which represents 8,5% of the sample, the largest educational status group is Bachelor's degree with 517 responses, which represents 62,5% of the sample, finally 207 participants reported that they have an associate degree, which represents 25% of the sample. Any of the participants were graduated from high school or lower. As expected, the majority of the participants had earned their Bachelor's degree. Table 19 provides a summary of participants' educational background.

Table 20: Sample Distribution by Title

	Frequency	Percent
Director	128	15,5
Ast. Director	68	8,2
Manager	251	30,4
Ast. Manager	48	5,8
Specialist	157	19,0
Ast. Specialist	84	10,2
Staff	91	11,0
Total	827	100,0

According to the results, a great portion of the participants are working at managerial level. The sum of directors and managers refers to 45,9% of the sample. Based on the seven categorizations of IT department titles, 128 participants indicated they work as director, representing 15,5% of the sample, 68 participants reported that they are assistant director, which represents 8,2% of the sample, almost one-third of the respondents rated themselves as manager, which refers to 251 participants and 30,4% of the sample, while 48 participants reported that they work as assistant manager, which refers to 5,8% of the sample, 157 professionals indicated they are specialist, which represents 19% of the sample, 84 participants reported that they work as assistant specialist, which refers to 10,2% of the sample, finally 91 responses are recorded as staff, which refers to 11% of the sample. Table 20 shows the frequency distribution in a summarized format.

Table 21: Sample Distribution by Organization Size They Work

	Frequency	Percent
1-9	27	3,3
10-49	116	14,0
50-249	263	31,8
250-499	282	34,1
500 and over	139	16,8
Total	827	100,0

Organization size was used as a demographic question. The sample consisted of Big Data professionals from all sizes of organizations. 27 participants reported that they work in a micro firm, representing 3,3% of the sample, 116 participants were from small firms which represent 14% of the sample, 263 professionals were from middle-sized enterprises, representing 31,8% of the sample, the largest response rate was from large enterprises with 282 professionals, representing 34,1% of the sample and finally 139 participants reported that they worked in very large enterprises, which represents 16,8% of the sample. Table 21 shows the frequency distribution in a summarized format.

Table 22: Number of IT Employees within The Workplace

	Frequency	Percent
1,00	88	10,6
2,00	291	35,2
3,00	256	31,0
4,00	137	16,6
5,00	55	6,7
Total	827	100,0

The number of IT employees was another demographic question regarding the enterprise that Big Data professionals worked. This was an open-ended question where the participants indicated the real employee number of the IT department. The answers are grouped and recorded into 5 variables since the maximum number of employees was 24. 88 participants indicated there were 1-5 IT workers, representing 10,6% of the sample, 291 participants reported there were 6-10 IT workers, which represents 35,2% of the sample, 256 responses indicated they have 11-15 IT workers, which represents 31% of the sample, 55 participants reported there were 16-20 IT workers, which refers to 16,6% of the sample and finally 55 participants reported that they worked with 21-25 IT workers, which represents 6,7% of the sample. Table 22 provides a summary of participants' responses.

Table 23: Number of Employees with Postgraduate Degree

	Frequency	Percent
,00	150	18,1
1,00	113	13,7
2,00	113	13,7
3,00	110	13,3
4,00	111	13,4
5,00	136	16,4

	Frequency	Percent
6,00	94	11,4
Total	827	100,0

We aimed to enlighten the educational structure of IT departments and asked the number of IT employees with a postgraduate degree. This was also scaled in ratio measure and the responses are gathered as real numbers. Recoding is not required since the maximum number of postgraduate employees was 6. 150 participants reported that there were no postgraduate employees in IT department, which refers to 18,1% of the sample and this was the largest response rate for this query, 113 participants indicated there was 1 employee with a postgraduate degree, which represents 13,7% of the sample, again 113 professionals reported there were 2 postgraduate employees in their department, representing 13,7% of the sample, 110 people indicated there were 3 postgraduate colleagues, which refers to 13,3% of the sample, 111 professionals reported there were 4 IT employees with postgraduate degree, which represents 13,4% of the sample, as the second largest response rate 136 participants indicated there were 5 postgraduate colleagues, representing 16,4% of the sample and finally 94 participants reported that there were 6 postgraduate member in IT department, which represents 11,4% of the sample. Table 23 presents the number of employees with a postgraduate degree within the IT departments in summarized form.

5.9. Exploratory Factor Analysis (EFA)

Factor analysis is conducted for the purpose of reducing a big set of variables or to scale items down to a slightly smaller, more manageable number of dimensions or factors. It does this by carefully "clumping" the items which are related to each other. This technique is used when developing scales and measurement.

There were 39 items in the questionnaire for the construct CSF as a whole since CSF was not measured before by dimensions an exploratory factor analysis is occupied to examine the clustering behavior of the items. There is a minimum sample requirement for factor analysis. As suggested by Tabachnick and Fidell (1996), based on the 5:1 ratio (5 cases for each item), the minimum sample size required for factor analysis for this study is 205 (39 (items) x5 (ratio)). Since the sample size after case deletion was 827, thus the sample was considered adequate to run factor analysis.

Table 24: KMO and Bartlett's Test

Construct	No. of items	KMO	Bartlett's
Team (TM)	7	0,859	,000
Success (S)	6	0,869	,000
Technology (TC)	7	0,849	,000
Project Management (PM)	7	0,849	,000
Project Definition (PD)	6	0,830	,000
Governance (G)	4	0,734	,000

As Tabachnick and Fidell (2001) cite Comrey and Lee's (1992) advice, as a rule of thumb, a bare minimum of 10 observations per variable is necessary to avoid computational difficulties. In our case, we exceed that observation but given the overall response rate, it was important to conduct the test of sampling adequacy. Two statistical tools were used to help assess the factorability of the data: Bartlett's t-test of sphericity (Bartlett, 1954) and the Kaiser- Meyer-Olkin (KMO) measure of sampling adequacy (Kaiser, 1970). This was done to ensure that the constructs exceeded a value of 0,5 or higher, indicating that the data was suitable for further testing (Kaiser, 1974). Furthermore, Bartlett's tests of Sphericity test to see if variables are unrelated and therefore unsuitable for structure detection. A significance level of less than 0.05 would indicate that a factor analysis would be useful.

The results of these two tests are reported in Table 24. The results in the table show that all the values of Bartlett's test are significant ($p\text{-value}<0,05$) and KMO results are more than 0,6 (min. 0,73) (Tabachnick and Fidell, 1996) for all the variables.

Principal component analysis (PCA) and Varimax rotation were used as the factor extraction and rotation method. This is because PCA is a psychometrically sound procedure and it is conceptually less complex than factor analysis. In addition to that, it bears numerous similarities to discriminant analysis (Field, 2009). Factor extraction was used to identify the minimum number of factors that can be used to best characterize the inter-relation among the set of variables. PCA is the most commonly used approach (Pallant, 2007).

Table 25: Rotated Component Matrix

Item	Component					
	1	2	3	4	5	6
TM3	,758					
TM4	,743					
TM2	,740					
TM1	,739					
TM5	,702					
TM6	,606					
TM7	,579					
S2		,779				
S3		,766				
S4		,750				
S5		,727				
S6		,638				
S1		,628				
TC4			,820			
TC5			,791			
TC3			,725			
TC6			,652			
TC2			,561			
TC1			,504			
TC7			,500			
PM4				,781		
PM3				,721		
PM5				,714		
PM7				,685		
PM2				,644		
PM6				,636		
PM1				,556		
PD3					,848	
PD4					,805	

Item	**Component**					
	1	2	3	4	5	6
PD5					,751	
PD2					,664	
PD6					,635	
PD1					,485	
G3						,776
G4						,714
G2						,698
G1						,456

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.
a. Rotation converged in 5 iterations.

Rotated component matrix helps to determine what the components are comprised of. It helps to focus on other analyses (reliability, correlation, and regression) on the variables that really create an impact. DiStefano et al. (2009) suggest an easy way to consider an item's relationship to the factor when creating a factor score, which is to include only items with loading values above a cut-off value in the computations. By doing so, researchers are only using "marker" items in the computation. However, the cut-off value to use is an arbitrary decision. In table, only the items with 0,45 or more loading are included. In table, all items are loaded in combinations relating to a component except for the item "It is critical to document during the project to aware of know-how loss." and "It is critical to access internal and external databases during the project.". These two items are excluded from EFA and further analyses because of their low loading values which is tested both by SPSS and SMART PLS within EFA, indicator reliability and discriminant validity tests. Other items are highly correlated to assigned constructs.

Table 26: Total Variance Explained

Component	Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Var.	Cmltv. %	Total	% of Var.	Cmltv. %
Team	10,755	29,068	29,068	4,256	11,501	11,501
Success	3,282	8,870	37,938	4,253	11,494	22,996
Technology	2,562	6,925	44,864	3,774	10,200	33,195
Project Management	2,115	5,716	50,579	3,718	10,049	43,244
Project Definition	1,603	4,333	54,913	3,263	8,818	52,062
Governance	1,376	3,720	58,632	2,431	6,570	58,632

Out of the 39 items in CSF, PCA revealed the presence of 6 components with eigenvalues exceeding 1, explaining 11,5%, 11,490%, 10,211.5%, 10,04%, 8,81% and 6,57% of variance, respectively. All items loaded on their related factor as expected except for two items, which had significant loadings (> 0,45) across multiple factors. Consequently, the two items were dropped from further analysis (Hair et al. 2010). The 6 components and 37 items were retained after 2 items were dropped due to factor loadings less than 0,45 and cross-loading with discrepancy less than 0,3 between the primary and secondary factor. The use of SMART PLS, on the other hand, specifies that cross loading is not allowed, which is more stringent compared to SPSS.

Table 27: Categories by Constructs

Category	Construct	Order	Item
Financial efficiency	Governance	G1	It is critical that the enterprise has the opportunity to invest in resources (technology, people, etc.) needed for the project.
Change management		G2	Ability to manage changes in technology, people, task and structure is critical within the scope of the project.

Category	Construct	Order	Item
Recruitment strategy		G3	It is critical to recruit appropriate and right people who fit the project needs.
Top management support		G4	Top management support is critical for the project.
Team leader skills	Team	TM1	It is critical that the project team leader has managerial abilities.
Multidisciplinary team		TM2	It is critical for the project team to involve employees from each relevant department.
Team skills		TM3	It is critical that people in the project team have analytical thinking skills.
		TM4	It is critical that people in the project team have the necessary technical skills.
Education status of the team		TM5	It is critical that the project team is trained or educated on Big Data.
Communication ability		TM6	It is critical that people in the team communicate with each other in a healthy way.
		TM7	Healthy interaction between project team end users is critical.
Technology infrastructure	Technology	TC1	It is critical that the enterprise has a flexible IT infrastructure.

Category	Construct	Order	Item
Defining technology strategy		TC2	Defining the strategy about which implementation and development tools will be used within the project is critical.
Data quality		TC3	It is critical that the data used in the project are qualified (complete, consistent, accurate, appropriate...).
Keep up with technology trends		TC4	The use of current analysis tools within the project is critical.
Technology infrastructure		TC5	The integration of old and new databases in the project is critical.
		TC6	It is critical that data management and auditing activities are performed hassle-free.
		TC7	It is critical to establish the IT infrastructure of the enterprise considering future needs.
Easy access to data sources		Q19	It is critical to access internal and external databases during the project.*
Task - technology - people balance	Project Management	PM1	Providing technology-task-people balance is critical in the project.
Allocation of resources		PM2	It is critical that the resources allocated to the project (human, technology, money, etc.) are properly distributed.
Team management		PM3	Proper job descriptions within the scope of the project are critical.
Project schedule		PM4	It is critical that the project schedule is clear.

Category	Construct	Order	Item
		PM5	It is of critical importance that the project is progressively benefited in a short time (milestones are quickly accessible).
		PM6	It is critical that the project will be concluded and implemented in a short time.
Defining measurement of success		PM7	It is critical that the project has measurable or definable outputs.
Documentation		Q27	It is critical to document during the project to aware of know-how loss.*
Defining business needs and objectives	Project Definition	PD1	It is critical that the business problem that the project aims to find a solution is well defined.
		PD2	Accurate identification of the purpose and scope of the project is critical.
		PD3	It is critical that the project is parallel to the business objectives.
Suitability		PD4	It is critical that the project scope fits the needs of the enterprise.
Positioning Big Data within the enterprise		PD5	It is critical to determine the strategic position of Big Data in achieving the objectives of the business.
Compatibility with business processes		PD6	It is critical that Big Data plays an important role in business and decision-making processes.

Category	Construct	Order	Item
Reaching project goals	Success	S1	The project objectives were successfully reached.
Satisfaction of end users		S2	The project end users were satisfied.
Reaching project budget targets		S3	The project has achieved the budget target.
Reaching quality goals		S4	The project has reached its quality target.
Reaching project schedule		S5	The project was completed in accordance with the project schedule.
Perception of success		S6	I believe the project was successful.

*Excluded item.

The hierarchical model was then estimated using the PLS path weighting scheme, which is the recommended method for estimating hierarchical latent variables, especially when the measurement model contains formative constructs (Becker et al. 2012; Rigdon, 2014). The strength of the inter-correlation in the correlation matrix was also checked using the SPSS (Statistical Package for the Social Sciences) software. All of them were more than 0,3 (min. 0,45) as recommended by Tabachnick and Fidell (1996).

5.9.1. Preliminary Statistics

Before moving further to analyze the relationships using PLS-SEM, some preliminary analysis on the variables was carried out to examine several tendencies in the data. Table 28 provides the descriptive statistics for the main variables: Governance, Team, Technology, Project Definition, Project Management, and Success. The average values were calculated for each variable using the items finalized in the EFA analysis. The mean, standard deviation, skewness, and kurtosis values were calculated and these are presented in Table 28.

Table 28: Descriptive Statistics of Constructs

	Governance	Technology	Team	Project Management	Project Definition	Success
Mean	5,8954	5,8508	6,2987	5,0230	5,5266	5,9405
Std. Deviation	,80964	,75100	,61664	,91321	,81787	,79003
Skewness	-,797	-1,033	-1,859	-,205	-,694	-,897
Std. Error of Skewness	,085	,085	,085	,085	,085	,085
Kurtosis	,649	1,309	5,066	-,738	,393	1,192
Std. Error of Kurtosis	,170	,170	,170	,170	,170	,170

Results indicate that Team, with 6,92 mean score, is the highest rated variable in the research model. Big Data professionals agree that Team related CSFs are very important for Big Data project success. Governance recorded the second highest mean score with 5,89. Professionals rated group of Technology items as the third important variable with mean score 5,85. Project Definition scored 5,52 at mean values and Project Management scored 5,02 which is the lowest mean within the research model. Success variable was about the Big Data project, that the professionals rated the CSFs for. So this variable does not refer to a CSF, rather it defines the measure of Success.

Table 29: Descriptive Statistics of Items

	Mean	Standard Deviation
TM1	6.111	0.812
TM2	6.086	0.837
TM3	6.255	0.760
TM4	6.421	0.827
TM5	6.470	0.863
TM6	6.401	0.891
TM7	6.346	0.906
PD1	5.585	1.187
PD2	5.862	0.950
PD3	5.532	1.137

	Mean	Standard Deviation
PD4	5.712	1.075
PD5	5.620	1.155
PD6	4.848	1.335
PM1	5.657	1.070
PM2	5.000	1.306
PM3	4.518	1.267
PM4	4.620	1.333
PM5	5.168	1.307
PM6	5.112	1.415
PM7	5.086	1.267
S1	5.709	1.046
S2	6.164	0.889
S3	6.017	0.977
S4	6.116	0.891
S5	5.944	0.956
S6	5.693	1.091
TC1	6.160	0.947
TC2	6.019	0.939
TC3	5.393	1.094
TC4	5.666	1.041
TC5	5.663	1.068
TC6	6.051	0.956
TC7	6.004	0.884
G1	5.784	1.046
G2	5.890	1.135
G3	5.941	1.014
G4	5.967	0.967

Table 28 presents the means and standard deviations of 37 items which are grouped under six constructs. TM5 (It is critical that the project team is trained or educated on Big Data.) has the highest mean score with 6,470. This indicates that the participants most agreed on this item. Education is one of the most important issues for niche IT projects. Big Data projects are special featured projects which involves expertness and know-how. Accordingly, it is expected an education based item to rate the highest mean (Agreement) score. Second highest rated item is TM4 (It is critical that people in the project team have the necessary technical skills.) with 6,421, followed by TM6 (It is critical that people in the team communicate with each other in a healthy way.), TM7 (Healthy interaction between project team end users is critical.) and TM3 (It is critical that people in the project team have analytical thinking skills.) respectively.

The lowest rated item is PM3 (Proper job descriptions within the scope of the project are critical.) with 4,518 mean score, followed by PM4 (It is critical that the project schedule is clear.), PD6 (It is critical that Big Data plays an important role in business and decision-making processes.), PM2 (It is critical that the resources allocated to the project (human, technology, money, etc.) are properly distributed.) and PM7 (It is critical that the project has measurable or definable outputs.). This indicates that the participants are closer to neutral on Project Management related issues.

5.9.2. Common Method Variance Biasness

Common method variance (CMV) is "attributable to the measurement method rather than to the constructs the measures represent" (Podsakoff, MacKenzie, Lee and Podsakoff, 2003). It concerns with systematic measurement errors/variance that either inflate or deflate the relationship between constructs. As a result, the results yielded may lead to misleading conclusions. Researchers should check for the common method variance when the source of data for the independent and dependent variables is generated from a single respondent (Simonin, 1999). For this study, there is a need to check for common method variance biasness because the data for both independent and dependent variables are collected from the same person via self- reported questionnaire (Podsakoff et al., 2003).

Data collected may be subject to common method variance biasness when a self-reported questionnaire is used to measure all the variables in a study. Conway and Lance (2010, p. 328) pointed out that "common method bias inflates relationships between variables measured by self-reports". According to Podsakoff and Todor (1985, p. 65), "invariably, when self-report measures obtained from the same sample are utilized in research, concern over same-source bias or general method variance arises''. Campbell (1982, p.

692) further elaborates that “If there is no evident construct validity for the questionnaire measure or no variables that are measured independently of the questionnaire, I am biased against the study and believe that it contributes very little and many people share this bias”.

Fundamentally, there are two distinct approaches to control for method bias. The first way is known as procedural remedies; which are to carefully design the study’s procedures (before data collection). The second method is statistical remedies to statistically control for the effects of method biases after the data collection. With regards to procedural remedies, Mackenzie and Podsakoff (2012) mentioned an array of aspects that potentially increase method bias by decreasing respondents’ ability and motivation to respond accurately. The authors presented all the possible causes for CMV and solutions to those problems. This study considers carefully all the relevant suggestions in the development of the questionnaire. For example, the researcher ensures that the questions asked in the survey form are clear, easy to understand and are not double-barreled. These were tested and confirmed in the pre-test stage. The length of the questionnaire is also kept at 15 minutes to avoid respondents’ fatigue and to minimize the cognitive effort required to answer the questions. Lastly, respondents are told to answer truthfully and that the information they provided will be kept confidential so that social desirability answers can be reduced.

In order to determine whether the common method variance is a problematic issue, Harman one-factor test is one of the commonly used methods (Krishnan, Martin, and Noorderhaven, 2006; Scott and Bruce, 1994). Items from all of the constructs in a study are included into factor analysis by running un-rotated principal component analysis. If the result shows that a single latent factor accounted for the majority of the explained variance, which is more than 50%, it indicates common method variance is an issue for the study (Podsakoff and Organ, 1986). Another way to detect common method variance is by looking at the inter-construct correlations using the correlation matrix. Common method variance can be a serious concern when the inter-construct correlations greater than 0,90 (Bagozzi, Yi, and Phillips, 1991).

In view of this, the Harman single-factor test was performed to identify the extent of this biasness (Ramayah, Lee, and In, 2011). According to Podsakoff and Organ (1986), common method bias is a concern if a single latent factor could describe the majority of the explained variance. In this study, the unrotated factor analysis showed that the factor accounted for most of the variance in the endogenous variable was 29,06%; thus, the common method bias was not a serious threat. In comparison, Doty and Glick (1998) in their study found common method variance of 32% a bias, yet reiterated that it does not invalidate many research findings.

5.10. Research Model

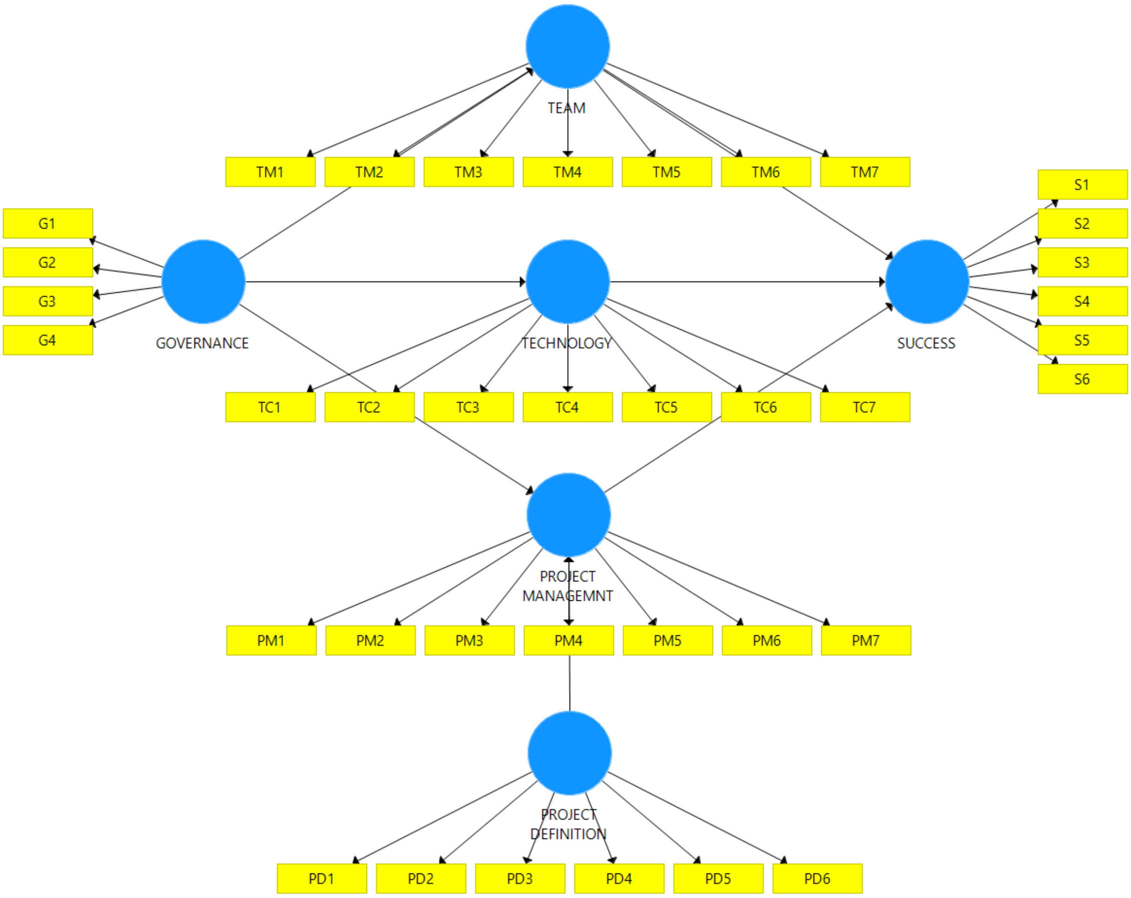

Figure 12: Research Model

In general, academic literature documents numerous articles dedicated to project success (de Wit, 1988; Mir and Pinnington, 2014; Müller and Jugdev, 2012). However, consensus on what defines project success (Mishra et al., 2011) is an evolving concept. As a result, the construct of project success continues to expand (Cserháti and Szabó, 2014; Davis, 2014) with additional clarity on what constitutes project success (Mir and Pinnington, 2014). This study reflected project success by five constructs namely Governance, Project

Definition, Project Management, Project Team, and Technology. The constructs emerged at the end of the EFA are defined in this section with regards to Big Data projects.

The use of the PLS model requires the test of both the measurement model as well as the structural model. In this study, Success was treated as a reflective measurement model while CSFs was treated as a formative measurement model. This is because CSFs are reflected by Governance, Team, Project Definition, Project Management, and Technology. In other words, the 5 CSFs are the factors that the project owners would consider when planning a Big Data project. The 5 CSFs did not form the Success context as there could be other factors that can be important to the Success "context" but were not used in calculating Success in this study. This is because studies related to Success are relatively small and in the exploratory form. This is the reason why exploratory factor analysis (EFA) was first carried out in this study; it was performed to gain a better understanding of the factors of Success. This implies that Success is subjected to changes and there could be more unexplored factors in Success. It should be noted that there is a minimum sample requirement when running PLS-SEM and the samples in this study were over those limits. Accordingly, this study enlightens a respectable percentage of Big Data project success but, because of the nature of the study, the results are limited to the samples' experiences.

5.10.1. Governance

Governance construct comprises of 4 items:

- It is critical that the enterprise has the opportunity to invest in resources (technology, people, etc.) needed for the project.
- Ability to manage changes in technology, people, task and structure is critical within the scope of the project.
- It is critical to recruit appropriate and right people who fit the project needs.
- Top management support is critical for the project.

These four items are both related to management abilities which are named as "Governance" in this study. Governance refers to the managerial perspective of organizational processes. In other words, it means business management. Items which are cumulated under this construct mainly based on top management support, recruitment and financial issues. Governance, management, and organizational capability constructs are used in IT-related research models or CSF studies with similar definitions (Gomez and Heeks, 2016; McAfee and Brynjolfsson, 2012; Saltz and Shamshurin, 2016; Evers, et al., 2014; Yeoh and Popovic, 2016; Gao et al., 2015; Nieder, 2016; Chen et al., 2016; Sadovskyi et al., 2014; Kamioka and Tapanainen, 2014; Cao and Duan, 2014; Wamba et al., 2016; Ji-Fan Ren, et al., 2016; Garmaki et al., 2016; Wang and Byrd, 2017; Dutta and Bose, 2015; Janssen et al., 2017; Koronios et al., 2014).

Defining the items under this construct enables deeper understanding. Top management support, which is one of the most well-known CSFs in IT success literature, refers to the willingness of top management to provide the necessary resources and authority/power for project success. Change management is another critical factor for most IT related projects. The item about change management reminds Leavitt's diamond which indicates CSFs of change management (Leavitt, 1965). Employee-related process is very crucial in terms of recruiting appropriate workforce and retention of valuable people in project related positions. Recruitment, selection, and training of the necessary personnel for the project team are also reflected by the Governance construct. The financial status of the enterprise is very important for project success. Big Data projects are expensive most of the time and financial strength to support emerging economic needs in terms of technology or people are needed for successful completion of Big Data projects.

5.10.2. Team

Team construct includes the seven items below:

- It is critical that the project team leader has managerial abilities.
- It is critical for the project team to involve employees from each relevant department.
- It is critical that people in the project team have analytical thinking skills.
- It is critical that people in the project team have the necessary technical skills.
- It is critical that the project team is trained or educated on Big Data.
- It is critical that people in the team communicate with each other in a healthy way.
- Healthy interaction between project team end users of the project output is critical.

These seven items are both related to project team which is named as "project team" or shortly "team" in this study. This construct reflects human-related issues in a project. Human capabilities are one of the most significant factors found in the CSF literature. It is named as people, human capability, project team or human/talent within the studies (Gomez and Heeks, 2016; McAfee and Brynjolfsson, 2012; Saltz and Shamshurin, 2016; McAfee and Brynjolfsson, 2012; Yeoh and Popovic, 2016; Gao et al., 2015; Nieder, 2016; Chen et al., 2016; Kamioka and Tapanainen, 2014; Cao and Duan, 2014; Wamba et al., 2016; Ji-Fan Ren, et al., 2016; Garmaki et al., 2016; Wang and Byrd, 2017; Kim and Park, 2016; Dutta and Bose, 2015; Koronois et al., 2014). The items grouped under Team refer to leadership skills, allocation of resources and technical and social skills. The importance of effective leadership skills in project management is frequently emphasized (Nguyen, et al., 2016). In the literature, the team is also linked to knowledge sharing Ghobadi, 2015; Chen and Hsieh, 2015) which is emphasized, in this study as communication with each other in a healthy way. This construct also reflects that the project team members had knowledge and skillset required for the organization had a data-driven culture. A data-driven culture is a one that where companies establish processes and operations to make it easy for employees to acquire the required information and are transparent about data access restrictions and governance methods. The structure of the team is also important (Martinez-Torres and Diaz-Fernandez; 2014; Ghobadi, 2015). The selected project team members should have high technical competence and expertise in project management processes and techniques. The project should provide appropriate technical training to the team, including training on the subject matter and Big Data processes. For Big Data projects it is very crucial that the project team worked in a facility with the supportive work environment, i.e. a healthy work culture that motivates employees to work for better project outcomes while keeping employee satisfaction as a primary concern. Also, the project should have a balanced team comprising an autonomous group of people with a variety of skills and perspectives to support achieving a common goal. This construct highlights that the project should strong communication focus and rigorous communication schedule, i.e. face-to-face and instant communication channels (between team members, between team and management, and between team and customers), daily stand-up meetings, build cycle meetings.

5.10.3. Project Management

Seven items grouped under Project Management construct are listed below:

- Providing technology-task-people balance is critical in the project.
- It is critical that the resources allocated to the project (human, technology, money, etc.) are properly distributed.
- Proper job descriptions within the scope of the project are critical.
- It is critical that the project schedule is clear.
- It is of critical importance that the project is progressively benefited in a short time (milestones are quickly accessible).
- It is critical that the project will be concluded and implemented in a short time.
- It is critical that the project has measurable or definable outputs.

The next area of review related to constructs includes Project Management. Project management as a methodology is not a separate function in many organizations today but is embedded in the management of the organization (Gopalasamy et al., 2013). In this sense, a project as a unique temporary endeavor could be better understood considering how the project and the project team interact and influence the overall organization and vice versa. Project Management is listed in CSF list of many IT projects and examined in several case studies (Saltz and Shamshurin, 2016; McAfee and Brynjolfsson, 2012; Yeoh and Popovic, 2016). The management of a project could be viewed in a broader context of the organization rather than just the project itself. Müller et al. (2014) emphasize for the governance of projects there is a need for a project to add to the overall value for the organization "a project is not an objective in itself but a means of achieving strategic change or future benefits" (p. 2). Additionally, project team interaction to the overall organization is important as positive interaction and adequate resourcing can influence the success of projects. Project Management refers to a methodological approach and structure to execute and manage projects (Pinto, 2014). Project Management is the process to provide oversight to individual projects also described as "the use of systems, structures of authority and processes to allocate resources and coordinate or control activity in a project" (Pinto, 2014, p. 8).

5.10.4. Project Definition

Project Definition construct consists of six items:

- It is critical that the business problem that the project aims to find a solution is well defined.
- Accurate identification of the purpose and scope of the project is critical.
- It is critical that the project scope fits the needs of the enterprise.
- It is critical that the project is parallel to the business objectives.
- It is critical to determine the strategic position of Big Data in achieving the objectives of the business.
- It is critical that Big Data plays an important role in business and decision-making processes.

Project Definition construct refers to the project scope and objectives were well-defined since the beginning of the project. The project objectives were business driven i.e. it captures the right business knowledge by communicating effectively with the stakeholders and within the realms of management, marketing, sales, operations, and other business affecting areas. Defining business cases are also very important. A business case provides a reference point before, during and after a project. The business case provides a value for what the objectives are and what they would deliver. Project definition can be evaluated as an original construct first introduced in this model in order to be tested statistically. Clear definition of a project is mentioned in several IT related project studies and this is found important. But, this study enhances the understanding of project definition and also strengthens the findings in the light of generated items and statistical analyses.

In any Big Data project, the team needs to have a deep understanding of project definition in order to plan an effective roadmap for successful completion of the project. They may have to take quick actions on unprecedented data management problems. It is necessary to have an action plan ready for such unforeseen problems. If not mitigated on time, these problems can become a huge threat later. The ever-increasing velocity and volume of data are two primary concerns. It is crucial to decide which data is important and needs storage, as opposed to which data is not useful in making business decisions. For this, it is necessary to have processes defined to determine which data has more business value and can be used in decision making. This step is very crucial as this data is going to drive the business decisions whose aftermath can be very crucial for the organization. Among the triple constraints, it is necessary to pay special attention to the scope of such projects and of course, the scope should be defined properly. It is necessary to have a well-defined goal. Metrics need to be defined to measure what degree of business value was derived from the project and how many objectives were obtained. Also, the expectations of the stakeholders should be clear and within the boundaries of the scope defined (Gao, Koronios and Selle, 2015).

5.10.5. Technology

Technology construct includes the seven items below:

- It is critical that the enterprise has a flexible IT infrastructure.
- Defining the strategy about which implementation and development tools will be used within the project is critical.
- It is critical that the data used in the project are qualified (complete, consistent, accurate, appropriate...).
- The use of current analysis tools within the project is critical.
- The integration of old and new databases in the project is critical.
- It is critical that data management and auditing activities are performed smoothly.
- It is critical to establish the IT infrastructure of the enterprise considering future needs.

Technology construct refers to the technology infrastructure of the enterprise and its ability to enhance technological maturity. This factor is related to technological tools, activities, and strategies. Technology is examined as a critical factor in IT projects in many studies. Researchers underline two main areas; the number of technology investments and technology readiness for the future. It is not always easy to change or improve the technology infrastructure of an enterprise. Change in tools or software enables improvement in the efficiency of work but on the other hand, these developments could change the business processes. Even it can be impossible to continue working during implementation and integration phases. This inflexible structure of technology issues makes it critical to implement the right and futurist technology strategy for enterprises.

The lack of proper solution designs or architecture for Big Data problems is among the prime technical problems. The technology that will be used should be customized according to the type of analysis which will be done via the project. Storage of data should be taken care of from the initial stages of the project. Data might be needed to be transformed into another form, to make it more structured, and make it a better fit for the business requirements. There is also a possibility of information loss during the process of transforming the unstructured data into a more structured format (Cuzzocrea, Song, and Davis, 2011). While merging the data from different sources, other concerns like security and privacy of data need to be taken into consideration as well. Access control mechanisms should be implemented to allow access to specific data to specific people depending on their role. Data spillage is an important concern especially when cloud-based platforms come into the picture. Storage, retrieval, and processing data on such cloud-based systems has a huge overhead especially when security comes into the picture (Gao, Koronios and Selle, 2015)

5.10.6. Success

Success construct is measured with the 6 items listed below:

- The project objectives were successfully reached.
- The project end users were satisfied.
- The project has achieved the budget target.
- The project has reached its quality target.
- The project was completed in accordance with the project schedule.
- I believe the project was successful.

Defining a project on the basis of satisfying the triple constraints of scope, schedule, and cost without looking at the overall business impact on the initial idea could lead to overall customer dissatisfaction. On the other hand, the new approach to project success, according to Shenhar and Dvir (2007) refers to business-related processes that are designed to deliver business results rather than a collection of project activities that have to be completed on time.

In general, academic literature documents numerous articles dedicated to project success (de Wit, 1988; Mir and Pinnington, 2014; Müller and Jugdev, 2012). However, consensus on what defines project success (Mishra et al., 2011) is an evolving concept. As a result, the construct of project success continues to expand (Cserháti and Szabó, 2014; Davis, 2014) with additional clarity on what constitutes project success (Mir and Pinnington, 2014). Ambiguity is further compounded by selecting appropriate instruments and measures viewed from various perspectives of the stakeholders trying to determine a complete and comprehensive meaning of success (Ika, 2009). The varying perspectives and meaning of project success from different stakeholders are expressed by Müller and Jugdev (2012) where project success continues to be "in the eyes of the beholder" (p. 763). This is consistent with McLeod et al. (2012) who discuss the perspectives of the stakeholders and associate these perspectives with a subjective philosophical approach for greater understanding of the meaning of those stakeholders' perceptions (p. 69). With the limited universal agreement of the definition of project success (Ika, 2009), various research methods and approaches are used in this study in order to discover what defines Big Data project success.

The items listed above are refined after a series of steps which are defined in previous chapters. According to these items Big Data project success refers to meet the expectations of the project owner(s). These are objectives, budget and quality targets and keeping up with the schedule.

It is hard to measure the success of a Big Data project. For example, if the effectiveness of a department or the enterprise increases after the implementation of the project this could be because of the project or another factor could also enhance effectiveness. This leads the researcher to measure perceived success instead of the real, statistically definable success criterion.

5.11. Hypotheses Development

The research conducted EFA to uncover the CSF constructs of Big Data project success. After the analysis, constructs emerged and hypotheses are developed to both validate and test the Big Data Success scale and Big Data Success Model.

Hypothesis 1:

Equally important, the performing organization or the governance board refers to those individuals who review the progress of the project and give the necessary approval or rejections for recommendations (Kerzner, 2006); the employees of the enterprise are the most directly involved in managing project activities (PMI, 2015). These employees are project managers and the project team members who are primarily concerned with the planning and implementation phases (Clements, 2008; Phillips, 2009; PMI, 2015) of the project life cycle. As Finch (2009) argued, the success of a project and the company depends upon the performing organization's ability to manage personnel effectively.

Decisions by project team members require individuals to commit to and are accountable in order their performance to improve, which will increase the performance of the overall project team (DeRond, 2012).

Employee issues regarding recruitment, selection, and training of the necessary personnel for the project team is generally controlled or

allowed by the governance team. Governance contains both financial and mental support for Big Data projects and also for project teams.

Accordingly, the first hypothesis is made;

H1: Governance has an effect on the project team.

Hypothesis 2:

Availability of the required technology and expertise to accomplish the specific technical action steps is under the control of business governance. The financial structure of the enterprise leads to qualified and adequate investment or weaknesses ending with unsuccessful projects. Ability to manage an enterprise requires farseeing and vision. Technology investments are one of the major criterion dealing with Big Data projects. Technological infrastructure and flexible solutions which are promising to meet new technological developments affects the success and sustainability of success of technology related projects. Big Data projects, which is also an IT project, require current analysis tools, database solutions, and technological modernity.

Thus, the second hypothesis is generated as;

H2: Governance has an effect on technology.

Hypothesis 3:

Business governance occupies a central position in every project. The PMI (2015) described project managers as the chief architects and executives in charge of overseeing the day-to-day project operations. Kerzner (2006) claimed that the project managers lead the project execution plan development from gathering the necessary resources (financial, human, equipment, etc.) to ensure deliverables across all project phases.

In project management, there are three basic organizational types: functional, project-based, and matrix (PMI, 2015). Each one has distinct hierarchical features that represent a largely internal view of the business (Axson, 2007) and set the level of authority, autonomy, and reporting structure for the project manager within the project (Phillips, 2004). According to Drucker (1954), people have to know and understand the organizational structure that they are supposed to work to avoid conflict in the reporting process.

The functional based organization is typically the type of structure used in industrial settings, especially in manufacturing settings, where external projects are rarely conducted (Gido and Clements, 2006; PMI, 2015). In this class of structure, the business managers have little power and autonomy; instead, they report directly to functional managers, and the project team is a part-time entity (Phillips, 2004). In contrast, the project-based structure is used by organizations that are solely into multiple projects at any one time and do not produce standard products (Gido and Clements, 2006; Kerzner, 2006). In project-based structures, according to Phillips, the managers enjoy a higher level of autonomy and responsibility for the projects, and they work on a full-time basis with the project team. Lastly, matrix structures are an amalgam of functional and project-based structures (Gido and Clements, 2006). The managers, as Phillips noted, have a full-time role and a reasonably high level of power.

The literature reveals a hypothesis regarding Governance and Project Management;

H3: Governance has an effect on Project Management.

Hypothesis 4:

Activities in the initiation, or the conceptualizing, phase mark the starting points of a project. Shenhar et al. (2001) asserted that the initiation phase of a project life cycle defines the strategic importance of the project to the enterprise. Other project experts have described the initiation stage of a project as the stage that defines and authorizes the project (Phillips, 2004; PMI, 2015); involves the identification of a need, problem, or opportunity; and can result in the customer requesting a proposal from a would-be performing organization (Gido and Clements, 2006).

This stage is characterized by the approval of a project charter. The power to launch the project or phase is given through a project definition (or project charter) (Phillips, 2004). Kerzner (2004) argued that the approval of the project definition is a generic process that often is omitted in organizations. Kerzner further stated that the project definition should be used to authorize work on the project; define the authority, responsibility, and accountability of the project team; and establish scope boundaries for the project. Other key effective practices in this phase of the project life cycle, according to Khang and Moe (2008), are to identify the potential beneficiaries and assess their development needs; align the development priorities of donors, the capacities of potential implementing agencies, and the development of needs; develop and evaluate project alternatives; and generate interest and support of key stakeholders.

Other researchers in the field of project management have defined project definition as the mechanism for translating strategic objectives into tactical actions (Aramo-Immonen and Vanharanta, 2009); an iterative process handled within the planning process group (Phillips, 2004); the art of asking, Who, What, When, Why, How Much, and How Long? and the determination of what needs to be done, by whom, and by when in order to fulfill one's assigned responsibility (Kerzner, 2006); preparation for the commitment of resources (Caughron and Mumford, 2008); determination of the details about the project (Wysocki, 2007); and the process of defining and maturing the project scope, developing the project management plan, and identifying and scheduling the project activities that occur within the project (PMI, 2015). Project definition is not a one-time approach; rather, it is iterative and interdependent (Gido and Clements, 2006; Kerzner, 2004; Newell, 2004; Phillips, 2004; PMI, 2015). Phillips claimed that project managers and their team return to the definition processes as often as needed throughout the project. As a result, experts in managing projects have suggested that the best approach is to allow the definition to go through the incremental or continuous process, otherwise known as progressive elaboration (Gido and Clements, 2006; Kerzner, 2006; Phillips, 2004; PMI, 2015) until the definition baseline has been produced.

According to Dhillon and Caldeira (2008) as well as Dvir (2005), a detailed definition and framework of a project is the only key to project success. Gold (1998) argued that organizational meanings of success may have no meaning without a mutual definition of constructing meaning. Defining the project is like a roadmap: it is a routine that describes the way that tasks are organized (Cicmil and Hodgson, 2006); a repeatable set of actions a team decides to perform on a regular basis to ensure that something is done in a certain way (Wearne, 2008); and a collection of activities that create value for a customer, a transformation of input/s into output/s (Angelides, 1999).

Successful project management practices on their own are not adequate to produce and deliver the desired products or services promptly and at minimal cost (Angelides, 1999).

Angelides argued that these practices must be integrated into the working framework of proven processes. Project success is how well the project definition meets the requirements of the end customer (Wysocki, 2007). A project definition that is understood promotes the teams' decision-making capabilities and aligns project management with the business strategy (Milosevic and Srivannaboon, 2006). Olsson (2006) claimed that organizations that re-engineer their business definitions gain sustainable competitive advantage.

In the light of these findings from the literature the researcher hypothesized;

H4: Project Definition has an effect on Project Management.

Hypothesis 5:

Another important stakeholder in the project is the project team. As Adam (2009) asserted, building high-quality teams does not happen by accident. It often needs to be encouraged by a determined, goal-orientated involvement that fosters greater self-awareness. Project teams refer to people who are working alongside project managers to deliver the actual work (Huemann, 2010); a group of interdependent individuals working cooperatively to achieve the project objective (Gido and Clements, 2006); a collection of individuals who will work together to ensure the success of the project (Phillips, 2004); the group that is performing the work of the project (PMI, 2015); or the group of people working towards a common objective (Dvir, 2005) to achieve success.

According to Axson (2007), ensuring effectiveness requires not only redesigned processes and new technologies but also appropriate skilled and trained practitioners who will make the right decisions; integrate, implement and transform data and information into knowledge (Heffner and Sharif, 2008); and take steps to ensure the achievement of the project goals. The importance of the project team, therefore, should not be taken slightly in any given project. As O'Dell, Grayson, and Essaides (1998) claimed, with technology, organizational problems are half solved, but the other half is not technology; rather, it is people. Methodologies do not manage projects; people do (Kerzner, 2006). Drucker (1954) argued that all organizations say routinely that people are their greatest asset, yet fewer practices what they preach, let alone truly believe it.

In line with Drucker's claim, Morris and Pinto (2007) argued that many companies are focused on the management of capital assets without any real measurements to monitor and make the most of a company's biggest asset: its people. As a result, Pinto suggested that effective practices demand that people be valued, measured, and developed because they are dynamic assets that can increase in value with time; they represent the remaining assets of a business after everything else has been eliminated, and company and shareholder values often suffer when human capital is mismanaged.

In conclusion, we made the fifth hypothesis;

H5: Project team has an effect on success.

Hypothesis 6:

This is the era of e-business or no business (Garrett, 2007). Miles (2003) asserted that enhancing process effectiveness could not become a reality until the development of the Internet. E-business, according to Garrett, describes a technology-enabled business that focuses on the seamless integration of the key stakeholders, performing organizations, project managers, and team members. This assertion suggests that technology is the vehicle that drives business success and that organization without technology face a higher risk of business failure. According to Schachter (2004), technology-related projects sometimes fail to meet project deadlines. Thus, if performing organizations are going to be successful, meet project deadlines, and ultimately satisfy their customers, they need technological tools (Thomas and Fernández, 2008).

Effective practices that drive higher returns and product superiority require the integration of technological tools and techniques (Besner and Hobbs, 2008) within the project life cycle. According to Spender (1996), technology is the master tool that shapes the systematic aspects of organizational systems. However, Starns and Odom (2006) asserted that managerial ability to identify the best mix of technologies through a combination of using current technologies, upgrading existing technologies where appropriate, and acquiring new technologies when required will be the principal factor in achieving project success.

The literature shed light on next hypothesis;

H6: Technology has an effect on success.

Hypothesis 7:

If there is no plan, there is no control (Hutka, 2009). As Dai and Wells (2004) asserted, project failure rates remain high, despite the advantages of project management methodology. As a result, planning techniques have received enormous attention (AramoImmonen and Vanharanta, 2009; Kerzner, 2006) based on the need for the appropriate control and management of large-scale projects (Caughron and Mumford, 2008; Dvir, 2005) to curb this failure rate.

Project managers also are responsible for managing the stakeholders' expectations (Gido and Clements, 2006; Hedeman et al., 2005; Phillips, 2009; PMI, 2015). Wood (2008) asserted that a well-informed and experienced project manager is an asset in terms of minimizing costs by utilizing the best practices suitable for the project, improving quality by reducing delays by preordering materials and equipment, and reducing risk through ongoing reviews and documentation. As a result, project managers must communicate unforeseen developments accurately to the team members and also listen to their suggestions (Hutka, 2009; Shenhar et al., 2007).

Successful project management contributes to the implementation of innovative ideas and influences the creative problem-solving process at much earlier stages of project development (Caughron and Mumford, 2008); enables accurate cost estimates to be produced; acts as an early warning system and keeps the project team focused (Gelbard and Carmeli, 2009); and reduces risks and the time required to complete the project (Söderholm, 2008). Successful management can help in the development of strategic information for customers to address risk and decide whether to commit resources to maximize the likelihood of a successful project (Griffith, Gibson, Hamilton, Tortora, and Wilson, 1999).

Project management plays a crucial role in the initial steps of the project. During the planning phase, the project managers and their teams meet, except when the project is virtual, to effectively plan their execution of the project. The activities entail planning the scope, cost, schedule, risks, quality, communication, human resources, contract, and procurement (PMI, 2015). Planning these aforementioned knowledge area perspectives requires the completion of a work breakdown structure to define the work necessary to produce the deliverables (Aramo-Immonen and Vanharanta, 2009).

Accordingly, the next hypothesis is made;

H7: Project Management has an effect on Success.

5.12. PLS Measurement Analysis

With regards to PLS, a number of analyses need to be performed. This includes the reflective measurement model analysis, formative measurement model analysis, and structural model analysis.

PLS-SEM relies on a nonparametric bootstrap procedure to test coefficients for their significance (Hair et al., 2010). In bootstrapping, large subsamples are taken from the original sample with replacement, i.e. every time an observation is randomly selected from the population; it is replaced back into the population before the next observation is selected. Therefore, the observation is always drawn from a population that always comprises all the same elements. A high number of bootstrap samples are preferred, with the minimum number of the bootstrap sample being at least equal to the number of valid observations in the dataset. According to Chin (1998), the bootstrapping size should have at least the total number of valid cases in the dataset. Hair et al., (2014) suggested 5000 bootstrap sample size as a rule of thumbs. With 827 valid data, 5000 bootstrap sample size is deemed to be sufficient to obtain consist results for this research. To validate all the hypothesized relationships, the direct effects were evaluated with the cut-off lower limit value of 0,05 for a regression coefficient, at the significance level of 5%.

Based on SEM literature, there are two broad types of measurement specification that researchers must realize when developing constructs – reflective and formative measurement models. Determining the type of measurement model is imperative because it will influence the subsequent evaluation process of the measurement models. Both reflective and formative measurement models have their own evaluation processes.

For reflective measurement models, the measures (items) represent the effects of an underlying construct, meaning the construct give meanings to the measurement of indicator variables, with the arrows pointing from the latent constructs to the reflective indicators (Hair et al., 2014). Because reflective indicators are assumed to contribute the same conceptual domain of the construct, reflective indicators are expected be closely related (highly correlated) to each other. Hence, individual items are supposed to be interchangeable. The removal of any item should not alter the conceptual domain of the construct, as long as adequate reliability is achieved.

For formative measurement models, the causality is from the indicators to its construct (Diamantopoulos, 2011). With multiple formative indicators, the direction of the arrows is from the indicators to the construct. Each item is not interchangeable because they capture a specific aspect of the construct's domain. The meaning of the construct is determined by a set of specific indicators. As a result, it indicates that removal of any item will potentially change the nature of the construct. Unlike reflective measurement approach, formative indicators should not be highly correlated because it may cause the problem of collinearity. As a result, the weights linking between formative indicators and its construct will become unstable and non-significant.

However, in some circumstances where researchers hope to operationalize the constructs at higher levels of abstractions, PLS-SEM allows higher-order models or hierarchical component models (HCMs). Conceptually it is to combine all the information of several lower-order constructs/dimensions (LOCs) into a latent construct (Lohmoller, 1989). There are four main types of HCMs as shown in Figure 13.

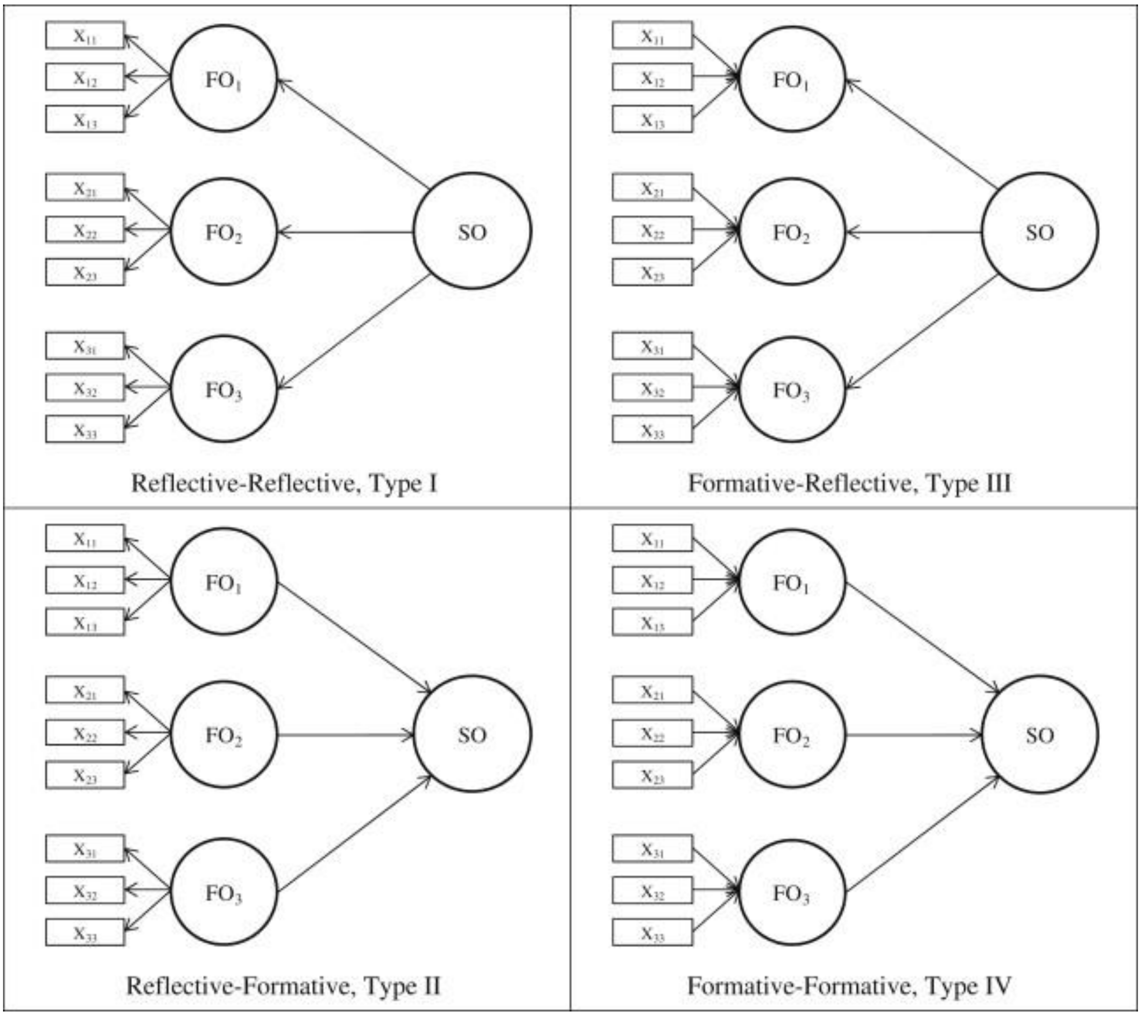

Figure 13: Hierarchical Latent Variable Models

(Becker, Klein and Wetzels, 2012)

There are two ways to establish the higher-order constructs/dimensions (HOCs) measurement models (Hair et al., 2014). The first way is to assign all the indicators from the LOCs to the HOC in the form of repeated indicators approaches. The second way is the two-stage approach involving two simple steps (this research is using this approach). Firstly, the repeated indicators approach is applied to obtain the latent variable scores for the LOCs. Then, the latent variable scores for the LOCs serve as the indicators for the HOC. It is important to note the same evaluation criteria of measurement models apply to the HOC.

The present study's quantitative research question inquires about causation; the research project is exploratory and the theoretical drive is quantitative. On the quantitative part of this study, Partial Least Squared (PLS) method was employed as the statistical tool to analyze the data. PLS-SEM analysis was used to analyze the moderating effect of the model path. SPSS software was required for basic statistical analysis and used in complimenting for Exploratory Data Analysis (EDA) and Exploratory Factor Analysis (EFA) while SMART PLS was the main software used for the structural model analysis.

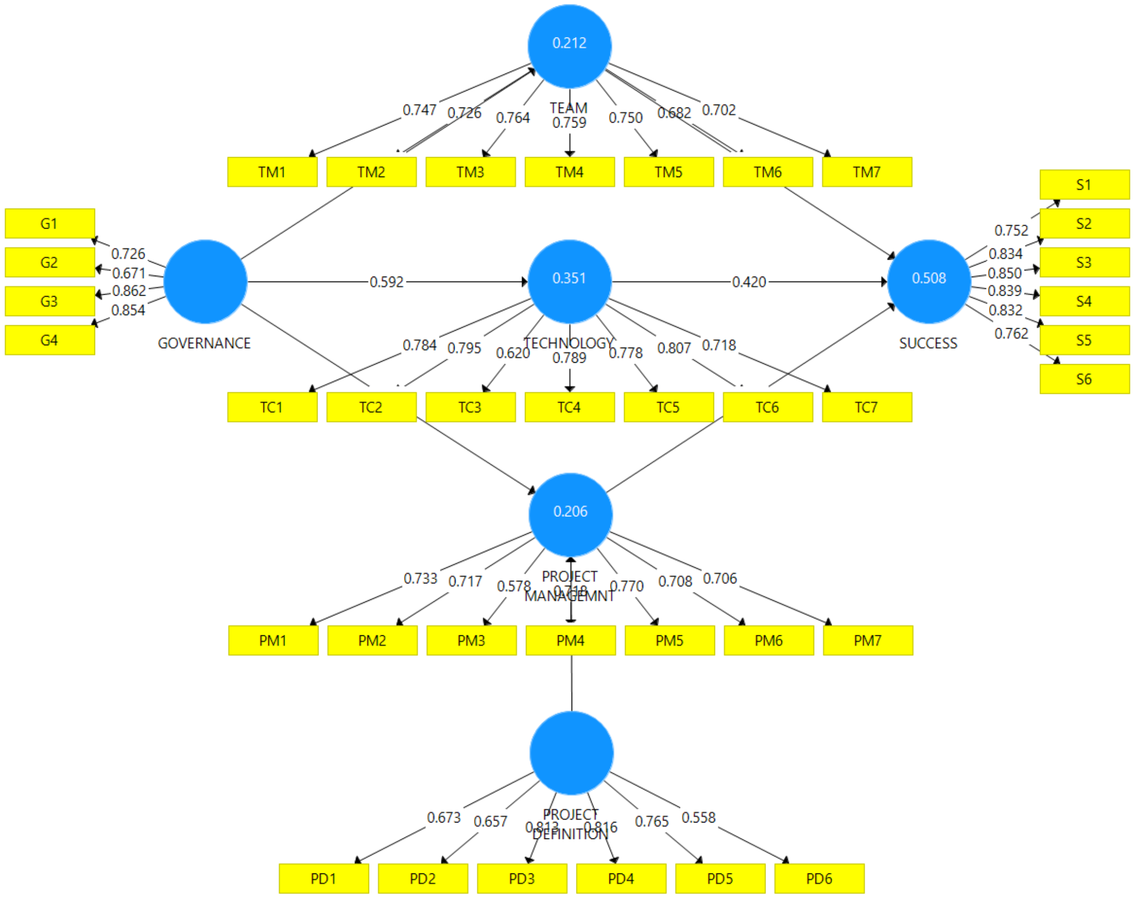

Figure 14: Research Model (validated)

5.12.1. Reflective Measurement Model

For reflective measurement models, it is important to assess the internal consistency of the items for the construct so that they are reliable. In addition, reflective construct should be uni-dimensional, which means each item of the construct is measuring the same conceptual domain of the construct. There are four important criteria to be assessed for reflective measurement model; namely internal consistency (assessment using composite reliability), indicator reliability (outer loadings), convergent validity (AVE) and discriminant validity (Fornell-Larcker and cross-loadings) so that both the validity and reliability of constructs can be achieved.

The evaluation is different for formative measurement models. The structural model estimates are not examined until the reliability and validity of the construct have been established (Hair et al., 2013). If the assessment of the reflective and formative measurement models provides evidence of the measure's quality, the structural model estimates are then evaluated. Hence, the primary evaluation criteria for PLS-SEM results are the coefficient of determination (R^2 value) as well as the level of significance of path coefficient.

5.12.1.1. Internal Consistency

The traditional criterion for internal consistency or the average correlation of items in a survey instrument is Cronbach's alpha, which is used to gauge its dependability (Santos, 1999). However, Cronbach's alpha undertakes that all indicators are equally reliable (all the indicators have equal outer loadings on the construct) while PLS-SEM selects the indicators according to their separate reliability. Cronbach's alpha is also sensitive to the number of items and generally underestimates the internal consistency reliability.

Cronbach's alpha: $\alpha = \frac{N}{N-1} \times 1 - (\frac{\sum_{i=1}^{N} \sigma_i^2}{\sigma_t^2})$ – Formula (A)

Where N = number of indicators assigned to a factor

σ_i^2 = The variance of indicator i

σ_t^2 = The variance of the sum of all assigned indicators' scores

Table 30: Construct Reliability and Validity

	Cronbach's Alpha	Composite Reliability
Governance	0,786	0,862
Project Definition	0,813	0,864
Project Management	0,837	0,874
Success	0,896	0,921
Team	0,856	0,890
Technology	0,876	0,904

Due to the limitations of Cronbach's alpha, composite reliability (also known as Dhillon-Goldstein Rho) is more appropriate to be used for PLS-SEM as it takes into account the different outer loadings of the indicator variables. The composite reliability refers to how well of a construct is explained by its own indicators. This measure provides a value which ranges between 0 and 1. Composite Reliability and Cronbach's alpha measures of 0,7 and above indicates sufficient convergence or internal consistency (Nunnally, 1978; Gefen, Straub and Boudreau, 2000). When reflective indicators do not achieve an outer loading greater than 0,4, those indicators should be removed immediately (Hulland, 1999). The results presented in table 30 indicate that the developed scale meets the quality criterion in respect to Nunnally's (1978) limits.

$$\text{Composite reliability} = \frac{(\sum \lambda_{ij})^2}{(\sum \lambda_{ij})^2 + \sum_i \ var(\epsilon_{ij})} \text{ – Formula (B)}$$

Where λ_i = loadings of indicator i of a latent variable

ϵ_i = measurement error of indicator i

j = flow index across all reflective measurement model

5.12.1.2. Indicator Reliability

Indicator reliability check was also carried out. Indicator reliability refers to the outer loading which indicates the proportion of indicator variance that is explained by the latent variable. It ranges between 0 and 1.

Table 31: Outer Loadings

	Management	Project Def.	Project Mng.	Success	Team	Technology
G1	0.726					
G2	0.671					
G3	0.862					
G4	0.854					
PD1		0.673				
PD2		0.657				
PD3		0.813				
PD4		0.816				
PD5		0.765				
PD6		0.558				
PM1			0.733			
PM2			0.717			
PM3			0.578			
PM4			0.718			
PM5			0.770			
PM6			0.708			
PM7			0.706			
S1				0.752		

S2				0.834		
S3				0.850		
S4				0.839		
S5				0.832		
S6				0.762		
TM1					0.747	
TM2					0.726	
TM3					0.764	
TM4					0.759	
TM5					0.750	
TM6					0.682	
TM7					0.702	
TC1						0.784
TC2						0.795
TC3						0.620
TC4						0.789
TC5						0.778
TC6						0.807
TC7						0.718

Reflective indicators with loadings that are less than 0,4 need to be removed (Hulland, 1999, p. 198). Indicators with outer loading between 0,4 and 0,7 are suggested for removal only when the deletion leads to a rise in the composite reliability and Average Variance Extracted (AVE) above the suggested threshold. If the AVE is above 0,5, the items with outer loadings in between 0,4 to 0,7 can be retained on its own construct. Reflective indicators with outer loading of higher than 0,7 should be retained in the model. The AVE criterion is described as the grand mean value of the squared loadings of the indicators related to the construct. Therefore, the AVE is equal to the communality of a construct. The removal of an indicator needs to be done carefully since the elimination may improve the reliability and discriminant validity but it may decrease the measurement's content validity.

5.12.1.3. Convergent Validity

For the Convergent Validity check, AVE was used. Convergent validity measures the extent to which a measure correlates positively with alternative measures of the same construct (Hair et al., 2014). The items that are indicators (measures) of a specific construct should converge or share a high proportion of variance. AVE value ranges between 0 and 1. Bagozzi and Yi (1988) and Fornell and Larcker (1981) suggest that variables with AVE exceeding 0,5 have adequate Convergent Validity. The calculation is presented in the formula below.

$$\text{Average Variance Extracted (AVE)} = \frac{\sum i\lambda_i^2}{\sum i\lambda_i^2 + \sum i var(\in i)} \text{ – Formula (C)}$$

Where λ_i^2 = square loadings of the indicator i of a latent variable

$var(\in i)$ = squared measurement error of indicator i

Table 32: Average Variance Extracted (AVE)

	Average Variance Extracted (AVE)
Governance	0,612
Project Definition	0,519
Project Management	0,501
Success	0,660
Team	0,538
Technology	0,575

The results represented in the table 32 indicate that each group of indicators meet or exceed the minimum accepted standards (Fornell and Larcker, 1981). Although one construct's AVE was just meet 0.5 (project management), it was considered to be acceptable. The discriminant validity criterion is considered to be fully satisfied if the square root of AVE of each construct is greater than the correlations between the construct and any other constructs (Fomell and Larcker, 1981; Chin, 1998).

Table 33: Latent Variable Correlations

	Governance	Project Definition	Project Management	Success	Team
Project Definition	0.288				
Project Management	0.407	0.311			
Success	0.526	0.336	0.459		

	Governance	Project Definition	Project Management	Success	Team
Team	0.461	0.195	0.245	0.513	
Technology	0.592	0.286	0.383	0.630	0.454

Table 33 presents the correlations between the constructs on the off-diagonal. The discriminant validity criterion is considered to be fully satisfied; however, Segars and Grover (1998) argue that this criterion "may be overly restrictive in some contexts." Accordingly, given that none of the off-diagonal correlations were higher than (almost equal to) their relevant diagonal elements in Table, there is a significant support for discriminant validity of the constructs. In addition, Bagozzi et al. (1991) suggest that the correlations between all pairs of constructs should be below 0.9 to ensure the distinctness of the constructs. All correlations on the off-diagonal in Table 33 were less than 0.9, indicating that the constructs are distinct. That means multiple measurements of the constructs are in agreement (Bagozzi, Davis, and Warshaw, 1992).

5.12.1.4. Discriminant Validity

Discriminant validity calculates the diversity of the constructs. Discriminant validity refers to the extent to which a construct is truly distinct from other constructs by empirical standards (Hair et al. 2014). A high discriminating validity is preferred as it indicates that a concept is specific and that some effects are ignored by other measures.

Table 34: Cross Loadings

	Governance	Project Definition	Project Management	Success	Team	Technology
G1	0.726	0.268	0.343	0.395	0.364	0.515
G2	0.671	0.213	0.247	0.329	0.274	0.344
G3	0.862	0.209	0.301	0.462	0.379	0.468
G4	0.854	0.210	0.364	0.445	0.405	0.498
PD1	0.273	0.673	0.303	0.292	0.184	0.236
PD2	0.191	0.657	0.174	0.249	0.208	0.232
PD3	0.153	0.813	0.189	0.193	0.067	0.138
PD4	0.239	0.816	0.244	0.308	0.190	0.266
PD5	0.230	0.765	0.226	0.233	0.143	0.215
PD6	0.060	0.558	0.127	0.086	-0.039	0.086

	Governance	Project Definition	Project Management	Success	Team	Technology
PM1	0.375	0.265	0.733	0.453	0.276	0.344
PM2	0.288	0.255	0.717	0.316	0.212	0.270
PM3	0.091	0.081	0.578	0.132	0.002	0.161
PM4	0.227	0.209	0.718	0.237	0.053	0.223
PM5	0.329	0.221	0.770	0.353	0.198	0.306
PM6	0.297	0.251	0.708	0.303	0.164	0.241
PM7	0.265	0.172	0.706	0.325	0.152	0.269
S1	0.393	0.272	0.330	0.752	0.386	0.507
S2	0.433	0.250	0.285	0.834	0.484	0.487
S3	0.437	0.290	0.396	0.850	0.439	0.512
S4	0.457	0.300	0.380	0.839	0.444	0.532
S5	0.430	0.255	0.433	0.832	0.418	0.518
S6	0.412	0.267	0.408	0.762	0.326	0.513
TC1	0.481	0.223	0.272	0.547	0.507	0.784
TC2	0.482	0.187	0.296	0.540	0.430	0.795
TC3	0.325	0.160	0.300	0.310	0.195	0.620
TC4	0.425	0.188	0.328	0.408	0.256	0.789
TC5	0.438	0.190	0.290	0.417	0.252	0.778
TC6	0.460	0.295	0.274	0.537	0.332	0.807
TC7	0.496	0.256	0.292	0.517	0.360	0.718
TM1	0.341	0.136	0.122	0.360	0.747	0.340
TM2	0.303	0.130	0.168	0.332	0.726	0.304
TM3	0.326	0.157	0.208	0.378	0.764	0.378
TM4	0.302	0.102	0.181	0.377	0.759	0.352
TM5	0.345	0.137	0.155	0.394	0.750	0.317
TM6	0.360	0.124	0.152	0.347	0.682	0.271
TM7	0.372	0.201	0.256	0.430	0.702	0.359

One of the methods to assess the discriminant validity is by examining cross-loadings of the indicators, known as Cross Loading Criterion (Chin, 1998a). A latent variable should explain better the variance of its own indicators than the variance of other latent variables so that the problem of multicollinearity is minimized (Chin, 1998a). As such, it also means that the loadings of an indicator must not be higher than in another construct rather than in its own assigned construct. As seen on the table, all indicators are located at the construct which they showed the highest loading.

Table 35: Fornell - Larcker Criterion

	Governance	Project Definition	Project Management	Success	Team	Technology
Governance	0,783					
Project Definition	0,289	0,720				
Project Management	0,407	0,309	0,706			
Success	0,526	0,339	0,459	0,813		
Team	0,461	0,202	0,245	0,513	0,733	
Technology	0,592	0,290	0,383	0,630	0,454	0,758

Fornell and Larcker's (1981) criterion is another approach to assess discriminant validity. It makes a comparison between the square root of AVE values with the latent variable correlations. The square root of each construct's AVE should be higher than its highest correlation with any other construct (Fornell and Larcker, 1981). To assess discriminant validity, latent construct's correlations matrices were used where the square roots of the AVEs along the diagonals are presented. Correlational statistics among constructs are presented in the lower left off-diagonal elements in the matrix. Discriminant validity is realized when the diagonal elements exceed the off-diagonal elements in the same row and column (Fornell and Larcker, 1981).

Table 36: Heterotrait-Monotrait Ratio (HTMT)

	Governance	Project Definition	Project Management	Success	Team
Project Definition	0,333				
Project Management	0,457	0,335			
Success	0,622	0,369	0,491		
Team	0,552	0,232	0,256	0,582	
Technology	0,696	0,318	0,432	0,697	0,505

Also, Heterotrait-Monotrait Ratio (HTMT) approach was introduced to better assess the discriminant validity of latent constructs. The new HTMT criteria is based on a comparison of the heterotrait-heteromethod correlations and the monotrait-heteromethod correlations. In order to establish discriminant validity, it was suggested that the acceptable value of HTMT must be lower than 0,9 (Henseler et al., 2015). However, a more rigid value was suggested that it must be lower than 0,85 in order to attain sufficient discriminant validity. Bootstrapping procedure needs to be performed in order to test whether the confidence interval contains the value of more than one. If the confidence interval is more than one, this indicates a lack of discriminant validity. Table 36 indicates that all HTMT values are lower than 0,85, thus sufficient discriminant validity is attained.

5.12.2. Formative Measurement Model

The statistical evaluation criteria for formative measurement models are different with reflective measurement models. In formative models, indicator contributes different conceptual domains of the formative constructs and therefore they do not necessarily correlate. For the formative measurement model, it is important to check for the collinearity among indicators as well as the significance and relevance of outer weights. Formative measurement models assessment procedure involves three crucial steps. The researcher needs to assess the convergent validity of formative measurement models, assess collinearity issues, and the significance and relevance of the formative indicators.

5.12.2.1. Convergent Validity

Convergent Validity is "the extent to which the measure correlates positively with other measures (indicators) for the same construct" (Hair et al., 2014. p. 121). The assessment of convergent validity of formative constructs is to determine the correlation between the formative measured construct with a reflective measure of the same construct, also known as redundancy analysis (Chin, 1998). For this research, CSFs are formative constructs. Redundancy assessment for convergent validity is not appropriate for reflective-formative constructs (perceived consequences) because it is made up of lower-order constructs (LOCs) representing different concepts (Hair et al., 2016).

5.12.2.2. Collinearity Issues

Formatively measured constructs indicators are not interchangeable; therefore, high correlations are not expected between indicators. When two formative indicators are highly correlated, this situation is known as collinearity; when more than two formative indicators involved, it is called as multicollinearity. This refers to the high correlations among indicators which can amplify the standard errors and thus reduces the ability to determine that the estimated weights significantly diverged from zero. The collinearity issues are problematic in PLS-SEM analysis since it depends on lesser sample sizes where standard errors are slightly higher due to sampling error. Additionally, high collinearity can result in the weights being wrongly projected as well as their signs being reversed (Hair et al., 2013).

A related measure of collinearity is the variance inflation factor (VIF). A Hair et al., (2014) pointed out that a VIF value of 5 and higher indicates a potential collinearity problem. Facing the problem with high levels of collinearity between formative indicators, one should consider deleting the items causing collinearity issue. Yet, the removal of problematic indicators is not as easy as the steps in reflective measurement models because it may alter the meaning of the formative constructs or insufficiently to capture the construct's content from a theoretical perspective. One should not continue to assess the significance and relevance of indicators if the collinearity problem is not solved.

Table 37: Outer VIF Values

Item	VIF	Item	VIF	Item	VIF
G1	1.305	**PM4**	1.953	**TC4**	2.603
G2	1.356	**PM5**	1.820	**TC5**	2.351
G3	2.452	**PM6**	1.575	**TC6**	2.174
G4	2.299	**PM7**	1.576	**TC7**	1.585
PD1	1.244	**S1**	1.954	**TM1**	1.952
PD2	1.568	**S2**	2.692	**TM2**	1.835
PD3	2.254	**S3**	2.815	**TM3**	1.934
PD4	2.273	**S4**	2.508	**TM4**	1.997
PD5	1.785	**S5**	2.358	**TM5**	1.929

Item	VIF	Item	VIF	Item	VIF
PD6	1.379	**S6**	1.966	**TM6**	1.538
PM1	1.533	**TC1**	2.609	**TM7**	1.520
PM2	1.606	**TC2**	2.585		
PM3	1.620	**TC3**	1.638		

The results in Table 37 show no collinearity issue as the outer VIF values are all below 5.0 (Hair, Ringle, and Sarstedt, 2011). This refers to the high correlations among indicators which can amplify the standard errors and thus reduce the ability to determine that the estimated weights significantly diverged from zero. The collinearity issue is problematic in PLS-SEM analysis since it depends on smaller sample sizes where standard errors are slightly higher due to sampling error. Additionally, high collinearity can result in the weights being wrongly projected as well as their signs being reversed (Hair et al., 2013).

5.12.2.3. Significance and Relevance of Formative Indicators

After checking the convergent validity and collinearity issue, the last step for the assessment of formative measurement models is to evaluate the contribution of the formative indicators by inspecting its outer weight (Hair et al., 2014).

Table 38: Outer Weights of Formative Constructs

	Governance	Project Definition	Project Management	Success	Team	Technology
G1	0.347					
G2	0.244					
G3	0.325					
G4	0.357					
PD1		0.331				
PD2		0.191				
PD3		0.207				
PD4		0.267				
PD5		0.247				
PD6		0.139				
PM1			0.284			

	Governance	Project Definition	Project Management	Success	Team	Technology
PM2			0.215			
PM3			0.079			
PM4			0.167			
PM5			0.233			
PM6			0.213			
PM7			0.200			
S1				0.195		
S2				0.199		
S3				0.211		
S4				0.214		
S5				0.214		
S6				0.198		
TM1					0.192	
TM2					0.174	
TM3					0.194	
TM4					0.187	
TM5					0.203	
TM6					0.194	
TM7					0.221	
TC1						0.212
TC2						0.210
TC3						0.130
TC4						0.171
TC5						0.175
TC6						0.205
TC7						0.208

If the result shows that outer loading is above 0,5, the indicator should be interpreted as having a strong one-to-one relationship with the dependent variable but it does not provide any explanatory power to the construct once other indicators have been added. If an indicator shows both insignificant outer weight and outer loading (<0,05), it is suggested to be removed because it might not theoretically support the conceptual domain.

Table 38 presents that all indicators' loadings are more that suggested minimal value. Accordingly, contributions of formative indicators are sufficient.

Table 39: Outer Loadings of Formative Constructs

	Governance	Project Definition	Project Management	Success	Team	Technology
G1	0.726					
G2	0.671					
G3	0.862					
G4	0.854					
PD1		0.673				
PD2		0.657				
PD3		0.813				
PD4		0.816				
PD5		0.765				
PD6		0.558				
PM1			0.733			
PM2			0.717			
PM3			0.578			
PM4			0.718			
PM5			0.770			
PM6			0.708			
PM7			0.706			
S1				0.752		
S2				0.834		
S3				0.850		
S4				0.839		
S5				0.832		
S6				0.762		
TM1					0.747	
TM2					0.726	

	Governance	Project Definition	Project Management	Success	Team	Technology
TM3					0.764	
TM4					0.759	
TM5					0.750	
TM6					0.682	
TM7					0.702	
TC1						0.784
TC2						0.795
TC3						0.620
TC4						0.789
TC5						0.778
TC6						0.807
TC7						0.718

For the formative measurement model, it is important to check for the collinearity among indicators as well as the significance and relevance of outer weights. The collinearity issue is problematic in PLS-SEM analysis since it depends on smaller sample sizes where standard errors are slightly higher due to sampling error. Additionally, high collinearity can result in the weights being wrongly projected as well as their signs being reversed (Hair et al., 2013).

In bootstrapping, samples of 5,000 were used as recommended by Hair et al. (2010). The result shows that none of the item's outer weight is significant. Thus, the formative indicator's outer loading would have to be analysed. The result shows that all formative indicators' outer loadings were above 0,5. The indicators with outer loadings above 0,5 were retained while those below 0,5 may be deleted. However, it was decided that these items should be retained because of the suggestion in the literature that these indicators are important measures. The values for the outer loadings of formative constructs are provided in Table 39. All indicators rated 0,5 or higher, minimum loading was 0,558 which is sufficient and means there is no collinearity problem among indicators.

5.12.3. Structural Model Validity

After constructs had been examined to be reliable and valid hypotheses, the seven hypotheses developed for this research were tested with structural equation modeling.

SEM employs a confirmatory approach rather than an exploratory approach to test the proposed model. Although similar to multiple regression analysis (MRA), SEM is a more powerful tool that "provides a powerful means of simultaneously assessing the quality of measurement and examining causal relationships among constructs" (Wang and Wang, 2012, p. 1).

Each hypothesis represents a specific relationship that must be specified in the structural model (Hair et al. 2010, p. 673). The key principles for measuring the structural model of PLS-SEM are: the significance of path coefficients, the level of R^2 (coefficient of determination) values the f^2 effect size (Cohen's f^2), the Q^2 predictive relevance and the q^2 effect size (Hair et al., 2014).

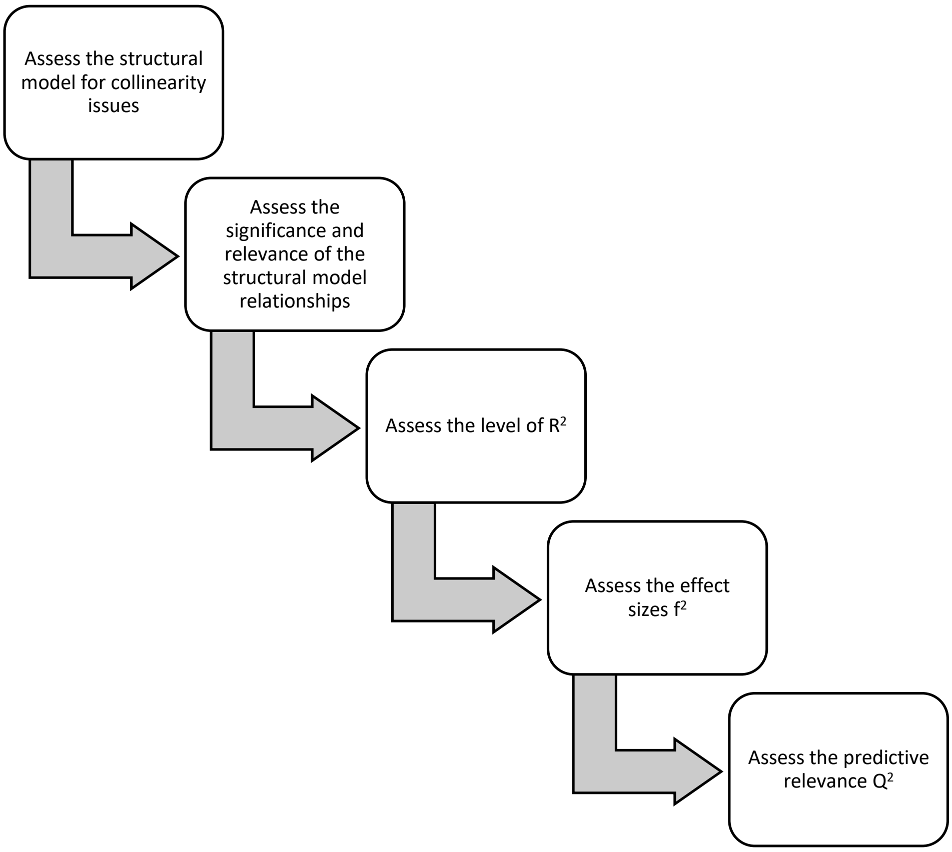

Figure 15: Structural Model Assessment Procedure

5.12.3.1. Collinearity

First, the structural model needs to be assessed for collinearity issues (aside from the formative measurement model). Collinearity problem occurs when there is a high correlation between the two constructs. If more than two constructs are involved, this situation refers to multi-collinearity. High levels of collinearity can be problematic because they have an impact on the estimation of weights and their statistical significance.

Table 40: Inner VIF Values

	Project Management	Success	Team	Technology
Governance	1.091		1.000	1.000
Project Definition	1.091			
Project Management		1.180		
Success				
Team		1.269		
Technology		1.397		

According to Hair, Ringle, and Sarstedt, (2010), a tolerance value of 0,20 or lower and VIF value of 5 and higher imply a potential collinearity problem exists.

It is suggested to eliminate the construct with collinearity issues, merge predictors into a single construct or create higher order constructs in order to treat the collinearity problems (Hair et al., 2013).

In order to validate the hypotheses, bootstrapping analysis is needed to test the direct effect of all the hypothesized relationship. The outcomes will determine whether the proposed relationship is significant. Chin (1998) recommended that bootstrapping sample size should be higher than the original sample so that random sampling errors can be mitigated from the bootstrapping procedure.

Inner VIF values indicate that the structural model has any collinearity problem among constructs. Maximum VIF value is 1,397 which means there is no need to eliminate any of the constructs.

5.12.3.2. Path Coefficients

The Path coefficient represents the hypothesized relationships among the constructs. The values range from -1 to +1 and are obtained after running the PLS-SEM algorithm. A value near |1| indicates a strong relationship while a value close to zero indicates a non-significant relationship. The value also indicates the direction of the relationship (positive or negative). However, to determine whether a coefficient is significant, it depends on the standard error obtained from bootstrapping.

Table 41: Path Coefficients

	Project Management	Success	Team	Technology
Governance	0,346		0,461	0,592
Project Definition	0,212			
Project Management		0,233		
Success				
Team		0,265		
Technology		0,420		

Table 41 depicts the path coefficients generated from the PLS analysis. As indicated in the research model in figure 14, all constructs have a significant positive impact on the constructs which are stated in the hypotheses. Hypothesis 1 stated a positive relationship between Governance and Team. Path coefficient of H1 is 0,461 which is the second highest value on the table and refers to a high positive impact. Hypothesis 2 stated a positive relationship between Governance and Technology. The highest impact seen on the research model was beet these two constructs with 0,592, accordingly, H2 was strongly supported. The path coefficient of H3 was 0,346, which is between Governance and Project Management. This value also indicates a statistically significant positive impact. Project Definition was expected to positively affect Project Management which is stated in H4. H4 was also supported with a path coefficient value of 0,212. H5 stated a positive relationship between Team and Success; this is also supported with a path coefficient value of 0,265. It is expected Technology to affect Success positively and that is stated in H6. The hypothesis is statistically significant with the third highest path coefficient value of 0,420. H7 indicated that there is a positive relationship between Project Management and Success. This is also accepted with 0,233 which refers to a statistically significant relationship.

5.12.3.3. Coefficients of Determination (R Square)

The most commonly used measure to evaluate the structural model is the coefficient of determination (R square). R square refers to how well the variance of a latent exogenous variable is explained by the total number endogenous latent variable expressed in a percentage (Chin, 1998). It measures the model's predictive accuracy and is measured as the squared correlation between a specific dependent construct and its predicted value. R square value ranges from 0 to 1 where higher levels indicate greater predictive accuracy. An R square value of 0,20 is considered high in disciplines such as consumer behavior while in marketing, 0,75, 0,50 and 0,25, as a rough rule of thumb, are described as substantial, moderate or weak, respectively (Hair, Ringle and Sarstedt, 2011; Henseler, Ringle and Sinkovics, 2009).

$$R^2 = \frac{SSR}{SS_{yy}} = 1 - \frac{SSE}{SS_{yy}} = 1 - \frac{SSE}{\sum y^2 - \frac{(\sum y)^2}{n}} \text{ – Formula (D)}$$

Table 42: Assessment of R-Square Values

Chin (1998)	Cohen (1988)	Assessment of R Square values
0,67	0,26	Substantial
0,33	0,13	Moderate
0,19	0,02	Weak

Different authors have suggested different assessment of R-square value as shown in Table 42. This research employs Chin's (1998) interpretation of R^2 as it is an enhanced version of Cohen's (1988). The calculation of R square is presented in the formula above.

Table 43: Coefficient of Determination (R-square)

	R Square	R Square Adjusted
Project Management	0,206	0,204
Success	0,508	0,506
Team	0,212	0,211
Technology	0,351	0,350

Table 43 summarizes the R-square scores for the independent variables, shown in the research model. R-square value for Technology is 0,351, which indicates 35% of the variance for Technology is driven by Management factor. On the other hand, the R-square value for Success is 0,518, which indicates 51,8% of success is accounted by the team, technology, project management, and project definition. The impact of management on the team is not that large with the R-square value of 0,212, accounting 21% of the variance from management. Besides, the R-square value for project management is only 0,20 accounting 20% of the variance from management and project definition. In structural equation models, path coefficients range greater than 0,1 is acceptable (Lohmöller, 1989). Accordingly, the variables and estimated relationships are left in the model.

Explanation rate of the research model is calculated as 51,8%, which is a very high explanation rate according to methodologically similar studies found in the literature. Table 44 represents explained variance of several well-known models.

Table 44: R-square Values of Reference Models

Theory/Model	Explained variance (R Square)	Reference
Theory of Reasoned Action (TRA)	0.36	Fishbein and Ajzen, 1975
Technology Acceptance Model - a (TAM2)	0.53	Venkatesh and Davis, 2000
- b (TAM- including gender)	0.52	Davis, 1989
Motivation Model (MM)	0.38	Davis, Bagozzi and Warshaw, 1992
Decomposed Theory of Planned Behavior (DTPB) - a TPB (including voluntariness)	0.36	Taylor and Todd, 1995b
- b TPB (including gender)	0.46	Ajzen, 1985
- c TPB (including age)	0.47	Ajzen, 1991
Combined Technology Acceptance Model and Theory of Planned Behavior (C-TAM-TPB)	0.39	Taylor and Todd, 1995a
Model of PC Utilization (MPCU)	0.47	Thompson et al., 1991
Innovation Diffusion Theory (IDT)	0.40	Rogers, 1995
Social Cognitive Theory (SCT)	0.36	Bandura, 1986

Theory/Model	Explained variance (R Square)	Reference
Unified Theory of Acceptance and Use of Technology (UTAUT)	0.69	Venkatesh et al., 2003
Modified TAM	0.44	Hu et al., 1999
IS Success Model	0,47	DeLone and McLean (1992)
Extended IS Success Model	0.41	Seddon, 1997
Amended Seddon Model	0.49	Rai, Lang and Welker, 2002

5.12.3.4. Effect Size (f^2)

The reason for determining the effect size of the predictor latent variables on the endogenous variables is to identify the strength of the particular predictor latent variable (Hair et al., 2014). The f^2 effect size refers to the changes in R^2 value when a specified independent construct is omitted from the model. The guidelines for measuring f^2 are that values of 0,02, 0,25 and 0,35 are small, medium and large effect (Cohen, 1988), respectively, of the exogenous latent variables.

$$\text{Effect size } f^2 = \frac{R^2_{included} - R^2_{excluded}}{1 - R^2_{included}} \text{ – Formula E}$$

Table 45: Assessment of f^2 Values

F2 Values	Assessment of F2 Values
0,02	Small
0,15	Medium
0,35	Large

The calculation of effect size is presented in formula and the assessment of f^2 is shown in Table 45.

Table 46: Effect Size (f^2)

	Project Management	Success	Team	Technology
Governance	0,138		0,269	0,540

Project Definition	0,052	
Project Management		0,094
Success		
Team		0,113
Technology		0,257

The independent variable that has the strongest effect size on Success is Technology (0,234) followed by Team (0,108) and Project Management (0,073). Furthermore, for Project Management, the predictor variable that has the greatest effect size is Governance (0,138). The strongest effect of the research model is seen between Governance and Technology (0,540), Governance – Team relationship has the second strongest effect size with 0,269.

5.12.3.5. Predictive Relevance (Q^2) and Effect Size (q^2)

In addition to evaluating R^2 values, the Stone-Geisser's Q^2 value (Geisser, 1974; Stone, 1974) should also be analyzed. This is one of the procedures in the structural model assessment. A measure of predictive validity is needed when employing SMART PLS for prediction purposes. It predicts the data points of indicators in the reflective measurement models of the endogenous constructs (Hair et al., 2013). Predictive validity of a complex model can be examined through blindfolding procedure by reproducing the observed values by the model itself and its parameter estimates (Tenenhaus et al., 2005). The blindfolding procedure can only be used for a reflective model. The Stone- Geisser criterion, Q^2 is a criterion to evaluate how well the omitted data are estimated by the model (Stone, 1974; Geisser, 1975). There are 2 modes to estimate the omitted data which are Cross-validated communality (represented as H^2 in the SMART PLS) or Cross-validated redundancy (represented as F^2 in SMART PLS). However, Hair et al., (2014) recommended using cross-validated redundancy approach as a measure of Q^2.

H^2 is where the missing values of the manifest data are estimated using the latent variables scores and factor loadings (Hair et al. 2014). F^2 is where the scores of the latent endogenous variables are estimated by the scores of latent exogenous variables and the weights in the measurement model (Hair et al. 2014). Then these newly estimated scores of latent exogenous variables are used to estimate the missing manifest variables scores. An important note is that the blindfolding procedure is only applicable to reflective endogenous constructs or endogenous single-item constructs. A model with a formative endogenous construct could not perform blindfolding procedure and thus predictive relevance criteria should be ignored.

Table 47: Assessment of Q2 Values

Q^2 Values	Assessment of Q^2 Values
0,02	Small
0,15	Medium
0,35	Large

Chin (1998) stated that Q^2 value that is larger than 0, indicates that the model has predictive relevance for the endogenous construct. According to Hanseler et al., (2009), the assessment of Q^2 is shown in Table 47.

Table 48: Q2 Values

Construct	Q^2 Values
Project Management	0,151
Success	0,353
Team	0,165
Technology	0,186

According to the results, the research model has large predictive relevance for the endogenous construct (Success), with a Q^2 value of 0,353. Also, other constructs which are affected by other exogenous constructs are also scored medium Q^2 values.

5.12.3.6. Model Fit

The model fit is the indices that determine the fit of the model. It is important to use a few indices to assess the model fitness because unlike CB-SEM, PLS-SEM does not optimize a unique global scalar function. The lack of global scalar function and the consequent lack of global goodness-of-fit measures are traditionally considered major drawbacks of PLS-SEM. The Standardized Root Mean Square Residual (SRMR) is currently the only approximate model fit criterion implemented for PLS path modeling (Henseler, Hubona, and Ray, 2016).

Table 49: Standardized Root Mean Square Residual (SRMR)

	Saturated Model	Estimated Model

SRMR	0,068	0,080
d_ULS	3.229	4.282
d_G1	0,904	0,922
d_G2	0,688	0,714
Chi-Square	3,336.208	3,383.397
NFI	0,787	0,784

The Normed Fit Index (NFI) is another useful approximate model fit criterion. However, NFI does not penalize for adding parameters; thus, it should be used with caution for model comparison. NFI's usage is also still rare.

Root Mean Square error correlation (RMStheta) is another approximate model fit criterion. Henseler et al. (2014) provide evidence that RMStheta can distinguish well-specified from ill-specified models. However, the threshold for RMStheta is yet to be determined.

Table 50: Root Mean Square error correlation (RMStheta)

RMStheta	0,109

The measurement model is the top-most concern since the questionnaire was designed based on the Delphi study. It must be noted that fit indices can be used as a guideline; however, it should be observed carefully. It is important not to move away from the original, theory testing purpose of structural equation modeling. There have been a lot of arguments regarding the 'rules of thumb' of the fit indices. It is highly controversial, with some experts urging for a complete abandonment of fit indices altogether (Barrett, 2007, Hair et al, 2013). Others are less certain of abandoning it and agree that adhering to the cut off values can lead to Type I error (Marsh, Hau, and Wen, 2004).

Measurement errors are discrepancies between the measured value and the estimated value of the object being measured (Field, 2009). Validity and reliability are the two properties of a measurement instrument that are essential for the trustworthiness of the data collection process and the resulting empirical findings. Validity refers to whether a measurement instrument measures what it is designed to measure and reliability is the consistency with which the instrument can be interpreted across different situations (Field, 2009).

After constructs had been examined to be reliable and valid hypotheses, the hypotheses developed for this research were tested with structural equation modeling. SEM employs a confirmatory approach rather than an exploratory approach to test the proposed model. Although similar to multiple regression analysis (MRA), SEM is a more powerful tool that ‘‘provides a powerful means of simultaneously assessing the quality of measurement and examining causal relationships among constructs’’ (Wang and Wang, 2012, p. 1).

DISCUSSION AND CONCLUSION

The problem in working with Big Data is the lack of improved analytical tools and platforms to solve multifaceted optimization problems, to support data formation of large masses of new kinds of data and relationships, and to discover and automate real-time and multifaceted decisions (Gandomi and Haider, 2015). Big data increases the need for refined statistics and analytical skills (Wixom et al., 2014). The era of Big Data is unique, primarily because the volume, velocity, and variety of data have changed (Yoon, Hoogduin, and Zhang, 2015). Data governance, privacy, and security challenges are producing a new level of concern from business leaders (Allen and Cervo, 2015). One of these concerns is the preparation of the IT department for the Big Data project. Big Data professionals should be aware of CSFs for successful completion of a Big Data project and finding these out is the primary reason for this research.

There has been limited empirical research on organizational factors that relate to Big Data (LaValle et al., 2011; Bean and Kiron, 2013). Even though there has been some empirical work on the technical, organizational, and individual factors related to Big Data adoption and success (Uğur and Turan, 2018; Al-Qirim et al., 2017), a gap exists in terms of understanding the critical success factors (CSFs), such as organizational size and top management support, that relate to Big Data project success. Previous studies have focused primarily on the technical and individual factors that relate to Big Data adoption. Sim (2014) acknowledged this gap and suggested that organizations should be aware of the important factors for Big Data success.

Unfortunately, a very limited amount of existing data, framework and variables exist concerning successful Big Data projects. It was, therefore, important to formulate a method that would allow us to collect data, review, analyze, deduce a model, formulate a theory and finally test the phenomenon statistically.

The best approach for such a study was mixed methods utilizing Constructivist Grounded Theory. Mixed methods allow for the integration of qualitative and quantitative data within a study to provide a more complete analysis of the research problem being investigated (Creswell and Plano Clark, 2011). It allows for, especially for an early concept, data to be built and further explored using a secondary method. Grounded theory allows the researcher to begin with the question, collect data, examine ideas and concepts, extract and categorize that data to use it to form the basis of a new theory. This new theory can then be applied and tested statistically. To successfully accomplish this, the approach for the study was fragmented into a three-part mixed methods study.

A qualitative section utilizing semi-structured interviews and Delphi study with experts in the field followed by a quantitative section to test relationships between core concepts derived from the qualitative section. First, conducting a qualitative study is suitable for the current research because the study is designed to investigate perceptions, experiences, and ideas (Ashby, Fryirs, and Howitt, 2015; Merriam, 2014). The qualitative technique is also useful for gathering a consensus opinion not found in the literature, an effort that would not be feasible with quantitative or mixed method approaches (Rees, Rapport, and Snooks, 2015). The qualitative portion of the study was done first, which allowed relationships to be tested later in a quantitative manner using statistical techniques. The knowledge gained through such a process allowed the quantitative section to be further insightful, concentrated and exploratory in nature.

Furthermore, it is important to note that this is also a correlational study (Kerlinger and Lee, 2000) as it is being conducted to determine the CSF relationships, as is, for successful Big Data projects. Standard strength and direction of relationships between variables are examined and predictions provided given the strength and conclusive nature of the variables within the study.

Prior to this current study, the CSFs that play a crucial role for successful completion of Big Data projects were slightly examined and the relationships between the CSFs were unknown as they are never been statistically tested and validated. The most complete information regarding the CSFs for a Big Data projects can be received from the Big Data professionals within those departments that have been involved in Big Data projects (Sivarajah et al., 2017). Accordingly, this study is conducted on 17 Big Data experts and 827 Big Data professionals. At the end of the study, five CSFs emerged and a statistically reliable and valid scale and a relational model is added to the literature.

This research may contribute to the IS success and Big Data literature by determining which success factors are critical for projects and examine if there is a statistical relationship between the factors. It is expected that most of the critical success factors will be consistent with previous studies within the IS success literature (Almajed and Mayhew, 2014; Palanisamy et al., 2010); but also there will be Big Data specific factors. Enlightening predictor success factors of Big Data projects are crucial because researchers have stated the need for determining the factors associated with Big Data (Sim, 2014).

Summary of Results and Findings

The main focus of this study was to examine the critical success factors that are essential for achieving success in Big Data projects. The purpose of this research was to build on the current diverse literature around Big Data by contributing discussion and data that allow common agreement on factors that influence successful Big Data projects.

This research is exploratory in nature. The best approach for such a study was mixed methods utilizing Constructivist Grounded Theory. To successfully accomplish this, the approach for the study was fragmented into a three-part mixed methods study. A qualitative section utilizing semi-structured interviews and Delphi study with experts in the field followed by a quantitative section to test relationships between core concepts derived from the qualitative section. The qualitative portion of the study was done first, which allowed relationships to be tested later in a quantitative manner using statistical techniques. The knowledge gained through such a process allowed the quantitative section to be further insightful, concentrated and exploratory in nature.

Qualitative Results and Findings

As described in the previous chapters, before the content of a new scale can be drafted, the researcher must define and understand the underlying construct, and articulate its connection to relevant existing theories, to aid to clarity in scale development (Clark and Watson, 1995; DeVellis, 2012).

Theories can drive, and also be the outcome of, the research process of coding qualitative data (Saldana, 2009). This study, as previously discussed, began theoretically, with qualitative data that will directly lead to scale items; Big Data Success Scale can then in consequent research be used to develop theory and test hypotheses.

Therefore, the first step of devising the Big Data Success Scale is to formulate a definition of the phenomena of “Big Data project success” and describe how this construct relates to other phenomena and their operationalization (DeVellis, 2012). A database of well-organized raw data forms a chain of evidence that allows the researcher to demonstrate that her interpretation of the data is firmly grounded in the data (Lazar, Feng, and

Hochheiser, 2010; Yin, 2003). In the present research, such a database starts with the results of a semi-structured interview study of the intended population. The definition of the “success” construct (also the dependent variable) then emerges from these data.

Before the Delphi study, semi-structured interviews with experts were conducted via phone (6 out of 17) and in most cases, face to face (11 out of 17) due to the sensitivity and confidential competitive advantage information regarding Big Data.

Answers were examined for conceptual commonalities and tallied. Six categories emerged from the data, these are: reaching project schedule, reaching project goals, reaching quality goals, reaching project budget targets, the satisfaction of end users and perception of success.

The researcher explored the following qualitative question with this research; *“What are the critical success factors that impact perceived project success in Big Data projects?”* In order to answer this question, after defining “Big Data project success” via semi-structured interviews, Delphi study is conducted in order to reveal CSFs of Big Data projects. The researcher's main intent for using the Delphi Technique was to reach a consensus among the experts. Semi-structured interviews and Delphi study are conducted on the same sample of experts. At the end of the first round, 83 statements were analyzed by the researcher, combining comparable comments and separating compound statements (Shinn et al., 2009). Through detailed thematic analysis, concepts and categories were developed, leading the researcher to identify 25 CSF statements representing CSFs of Big Data projects. These are: Financial efficiency, change management, recruitment strategy, top management support (added from the literature and served at the second round), team leader skills, multidisciplinary team, team skills, education status of team, communication ability, technology infrastructure, defining technology strategy, data quality, keeping up with technology trends, technology infrastructure, easy access to data sources, task - technology - people balance, allocation of resources, team management, project schedule, defining measurement of success, documentation, defining business needs and objectives, suitability, positioning Big Data within enterprise, compatibility with business processes. At the second round, seventeen experts rated these elements based on the importance on a scale of 0- 7 with zero (0) indicating “strongly disagree” and seven (7) indicating “strongly agree”. Both the median and interquartile range was calculated for each element listed. At the end of the second consensus was reached for all items with IQR $\leq$1.

Quantitative Results and Findings

Quantitative efforts constructed a CSF scale and a statistical research model in consideration of qualitative findings. The research also validated the CSF scale and relational CSF model statistically.

The researcher followed well-accepted procedures for the conceptual development of factor identification (Hair et al., 2010) and the scale development process (Churchill, 1979; Crocker and Algina, 1986; DeVellis, 2003; Gerbing and Anderson, 1988; Netemeyer, Bearden, and Sharma, 2003; Nunnally and Bernstein, 1994) which are found from a review of the current literature. This process involves construct definition, item generation, and purification, content validity, reliability and validity assessments. The process involved an inductive approach by relying on qualitative analysis to generate scale items to measure the constructs.

After the scale development process was ended and the pretest and piloting of the scale were completed, a private firm was hired by the researcher to conduct the survey. The survey administered through a CATI system. It included an IT workers sample. The use of survey method with CATI technique enabled the researcher to gather information nationally. The survey comprising of a total of 50 questions scaled with multiple choice and the Likert scale was conducted through telephone. A total number of 912 responses are recorded via CATI. After a rigorous filtering process, 827 usable responses left, which is sufficient for structural equation modeling.

Several demographic questions were included in the questionnaire regarding gender, age, education status, firm size, industry, IT and Big Data experience of participants etc. Industry statistics revealed that the finance industry was one of the greatest investors of Big Data technologies. Industry segmentation of the sample highlighted the big players of the Big Data market. IT companies may refer to outsource services about Big Data solutions. After the IT industry, finance with 13,2% and retail with 11,7% led the segmentation.

Factor analysis is conducted for the purpose of reducing a big set of variables or to scale items down to a slightly smaller, more manageable number of dimensions or factors. Out of the 39 items in CSF, PCA revealed the presence of 6 components with eigenvalues exceeding 1, explaining in total 58,632% of the variance.

Results indicate that Team, with 6,92 mean score, is the highest rated variable in the research model. Big Data professionals agree that Team related CSFs are very important for Big Data project success. Governance recorded the second highest mean score with 5,89. Professionals rated group of Technology items as the third important variable with mean score 5,85. Project Definition scored 5,52 at mean values and Project Management scored 5,02 which is the lowest mean within the research model.

The researcher explored the following quantitative question with this research; *"What are the relationships among the critical success factors?"*.

As part of this research question, after EFA is conducted, several hypotheses mentioning relations between the CSFs are examined. The relations between CSFs are visualized and tested in a relational model via SEM. The hypotheses and results are summarized in table 51.

Table 51: Hypothesis Results

Hypothesis		**Path Coefficient**	**Result**
H1	Governance has an effect on Project Team.	0,461	Strongly Supported
H2	Governance has an effect on Technology.	0,592	Strongly Supported
H3	Governance has an effect on Project Management.	0,346	Strongly Supported
H4	Project definition has an effect on Project Management.	0,212	Strongly Supported
H5	Team has an effect on Success.	0,265	Strongly Supported
H6	Technology has an effect on Success.	0,420	Strongly Supported
H7	Project Management has an effect on Success.	0,233	Strongly Supported

Hypothesis 1: Governance has an effect on Team.

Equally important, the performing organization or the governance board refers to those individuals who review the progress of the project and give the necessary approval or rejections for recommendations (Kerzner, 2006); the employees of the enterprise are the most directly involved in managing project activities (PMI, 2015). These employees are project managers and the project team members who are primarily concerned with the planning and implementation phases (Gido and Clements, 2006; Phillips, 2009; PMI, 2015) of the project life cycle. As Finch (2009) argued, the success of a project and the company depends upon the performing organization's ability to manage personnel effectively. Decisions by project team members require individuals to commit to and are accountable in order to improve their performance, which will increase the performance of the overall project team (DeRond, 2012). Employee issues regarding recruitment, selection, and training of the necessary personnel for the project team is generally controlled or allowed by the governance team. Governance contains both financial and mental support for Big Data projects and also for project teams (Yeoh and Popovic, 2016). In conclusion, it is expected Governance to have an effect on project team, this hypothesis is accepted and the relations between these two CSFs are statistically validated. Previous studies found in the literature also support the existence of this relation.

Hypothesis 2: Governance has an effect on Technology.

Availability of the required technology and expertise to accomplish the specific technical action steps is under the control of business governance (Gomez and Heeks, 2016; Yeoh and Popovic, 2016). The financial structure of the enterprise leads to qualified and adequate investment or weaknesses ending with unsuccessful projects. Ability to manage an enterprise requires farseeing and vision. Technology investments are one of the major criterion dealing with Big Data projects. Technological infrastructure and flexible solutions which are promising to meet new technological developments affect the success and sustainability of the success of technology-related projects (Wang and Byrd, 2017). A Big Data project, which is also an IT project, requires current analysis tools, database solutions, and technological modernity. Thus, it is suggested that Governance has an effect on technology and this relation is supported by empirical evidence as well as the previous literature.

Hypothesis 3: Governance has an effect on Project Management.

Business governance occupies a central position in every project. The PMI (2004) described project managers as the chief architects and executives in charge of overseeing the day-to-day project operations. Kerzner (2006) claimed that the project managers lead the project execution plan development from gathering the necessary resources (financial, human, equipment, etc.) to ensure deliverables across all project phases.

In project management, there are three basic organizational types: functional, project-based (projectized), and matrix (PMI, 2015). Each one has distinct hierarchical features that represent a largely internal view of the business (Axson, 2007) and set the level of authority, autonomy, and reporting structure for the project manager within the project (Phillips, 2004). According to Drucker (1954), people have to know and understand the organizational structure that they are supposed to work to avoid conflict in the reporting process.

The functional based organization is typically the type of structure used in industrial settings, especially in manufacturing settings, where external projects are rarely conducted (Gido and Clements, 2006; PMI, 2015). In this class of structure, the business managers have little power and autonomy; instead, they report directly to functional managers, and the project team is a part-time entity (Phillips, 2004). In contrast, the project-based structure is used by organizations that are solely into multiple projects at any one time and do not produce standard products (Gido and Clements, 2006; Kerzner, 2006). In project-based (projectized) structures, according to Phillips, the managers enjoy a higher level of autonomy and responsibility for the projects, and they work on a full-time basis with the project team. Lastly, matrix structures are an amalgam of functional and project-based structures (Gido and Clements, 2006). The managers, as Phillips noted, have a full-time role and a reasonably high level of power.

The literature reveals a hypothesis regarding Governance and Project Management; in other words, governance characteristics seem effective in Project Management. The related hypothesis is accepted in this study, evidence from the literature and findings of previous studies also supported this relationship.

Hypothesis 4: Project definition has an effect on Project Management.

Activities in the initiation, or the conceptualizing, phase mark the starting points of a project. Shenhar et al. (2007) asserted that the initiation phase of a project life cycle defines the strategic importance of the project to the enterprise. Other project experts have described the initiation stage of a project as the stage that defines and authorizes the project (Phillips, 2004; PMI, 2015); involves the identification of a need, problem, or opportunity; and can result in the customer requesting a proposal from a would-be performing organization (Gido and Clements, 2006).

This stage is characterized by the approval of a project charter. The power to launch the project or phase is given through a project definition (or project charter) (Phillips, 2004). Kerzner (2004) argued that the approval of the project definition is a generic process that often is omitted in organizations. Kerzner further stated that the project definition should be used to authorize work on the project; define the authority, responsibility, and accountability of the project team; and establish scope boundaries for the project. Other key effective practices in this phase of the project life cycle, according to Khang and Moe (2008), are to identify the potential beneficiaries and assess their development needs; align the development priorities of donors, the capacities of potential implementing agencies, and the development of needs; develop and evaluate project alternatives; and generate interest and support of key stakeholders.

Other researchers in the field of project management have defined project definition as the mechanism for translating strategic objectives into tactical actions (Aramo-Immonen and Vanharanta, 2009); an iterative process handled within the planning process group (Phillips, 2004); the art of asking, Who, What, When, Why, How Much, and How Long? and the determination of what needs to be done, by whom, and by when in order to fulfill one’s assigned responsibility (Kerzner, 2006); preparation for the commitment of resources (Caughron and Mumford, 2008); determination of the details about the project (Wysocki, 2007); and the process of defining and maturing the project scope, developing the project management plan, and identifying and scheduling the project activities that occur within the project (PMI, 2015). Project definition is not a one-time approach; rather, it is iterative and interdependent (Gido and Clements, 2006; Kerzner, 2004; Newell, 2002; Phillips, 2004; PMI, 2015). Phillips claimed that project managers and their team return to the definition processes as often as needed throughout the project. As a result, experts in managing projects have suggested that the best approach is to allow the definition to go through the incremental or continuous process, otherwise known as progressive elaboration (Gido and Clements, 2006; Kerzner, 2006; Phillips, 2004; PMI, 2015) until the definition baseline has been produced.

According to Dhillon and Caldeira (2008) as well as Dvir (2005), a detailed definition and framework of a project is the only key to project success. Gold (1998) argued that organizational meanings of success may have no meaning without a mutual definition of constructing meaning. Defining the project is like a roadmap: it is a routine that describes the way that tasks are organized (Cicmil and Hodgson, 2006); a repeatable set of actions a team decides to perform on a regular basis to ensure that something is done in a certain way (Wearne, 2008); and a collection of activities that create value for a customer, a transformation of input/s into output/s (Angelides, 1999).

Successful project management practices on their own are not adequate to produce and deliver the desired products or services promptly and at minimal cost (Angelides, 1999). Angelides argued that these practices must be integrated into the working framework of proven processes. Project success is how well the project definition meets the requirements of the end customer (Wysocki, 2007). A project definition that is understood promotes the teams' decision-making capabilities and aligns project management with the business strategy (Milosevic and Srivannaboon, 2006). Olsson (2006) claimed that organizations that re-engineer their business definitions gain sustainable competitive advantage.

Project definition is a brand new construct emerged in this study, especially with this scope and name. But the literature contains several pieces of evidence for the existence of such a construct and its relation with the project management. It is statistically proved that the project definition has an effect on project management and this construct is added to relevant literature for further use, test and validation.

Hypothesis 5: Project team has an effect on success.

Another important stakeholder in the project is the project team. As Adam (2009) asserted, building high-quality teams does not happen by accident. It often needs to be encouraged by a determined, goal-orientated involvement that fosters greater self-awareness. Project teams refer to people who are working alongside project managers to deliver the actual work (Huemann, 2010); a group of interdependent individuals working cooperatively to achieve the project objective (Gido and Clements, 2006); a collection of individuals who will work together to ensure the success of the project (Phillips, 2004); the group that is performing the work of the project (PMI, 2015); or the group of people working towards a common objective (Dvir, 2005) to achieve success.

According to Axson (2007), ensuring effectiveness requires not only redesigned processes and new technologies but also appropriate skilled and trained practitioners who will make the right decisions; integrate, implement and transform data and information into knowledge (Heffner and Sharif, 2008); and take steps to ensure the achievement of the project goals. The important of the project team, therefore, should not be taken slightly in any given project. As O'Dell, Grayson, and Essaides (1998) claimed, with technology, organizational problems are half solved, but the other half is not technology; rather, it is people. Methodologies do not manage projects; people do (Kerzner, 2006). Drucker (1954) argued that all organizations say routinely that people are their greatest asset, yet few practices what they preach, let alone truly believe it.

In line with Drucker's claim, Pinto (2007) argued that many companies are focused on the management of capital assets without any real measurements to monitor and make the most of a company's biggest asset: its people. As a result, Pinto suggested that effective practices demand that people be valued, measured, and developed because they are dynamic assets that can increase in value with time; they represent the remaining assets of a business after everything else has been eliminated, and company and shareholder values often suffer when human capital is mismanaged.

The project team is one of the most discussed issues within project management literature. It is expected to exist in a relationship between project team and success. Our predictions are statistically supported and strengthened by the cases found in the literature.

Hypothesis 6: Technology has an effect on success.

This is the era of e-business or no business (Garrett, 2007). Miles (2003) asserted that enhancing process effectiveness could not become a reality until the development of the Internet. E-business, according to Garrett, describes a technology-enabled business that focuses on the seamless integration of the key stakeholders, performing organizations, project managers, and team members. This assertion suggests that technology is the vehicle that drives business success and that organization without technology face a higher risk of business failure. According to Schachter (2004), technology-related projects sometimes fail to meet project deadlines. Thus, if performing organizations are going to be successful, meet project deadlines, and ultimately satisfy their customers, they need technological tools (Thomas and Fernández, 2008).

Effective practices that drive higher returns and product superiority require the integration of technological tools and techniques (Besner and Hobbs, 2008) within the project life cycle. According to Spender (1996), technology is the master tool that shapes the systematic aspects of organizational systems. However, Starns and Odom (2006) asserted that managerial ability to identify the best mix of technologies through a combination of using current technologies, upgrading existing technologies where appropriate, and acquiring new technologies when required will be the principal factor in achieving project success. Thus, it is suggested that technology has an effect on success and this relation is supported by empirical evidence as well as the previous literature.

Hypothesis 7: Project management has an effect on success.

If there is no plan, there is no control (Hutka, 2009). As Dai and Wells (2004) asserted, project failure rates remain high, despite the advantages of project management methodology. As a result, planning techniques have received enormous attention (AramoImmonen and Vanharanta, 2009; Kerzner, 2006) based on the need for the appropriate control and management of large-scale projects (Caughron and Mumford, 2008; Dvir, 2005) to curb this failure rate.

Project managers also are responsible for managing the stakeholders' expectations (Gido and Clements, 2006; Hedeman et al., 2005; Phillips, 2009; PMI, 2015). Wood (2008) asserted that a well-informed and experienced project manager is an asset in terms of minimizing costs by utilizing the best practices suitable for the project, improving quality by reducing delays by preordering materials and equipment, and reducing risk through ongoing reviews and documentation. As a result, project managers must communicate unforeseen developments accurately to the team members and also listen to their suggestions (Hutka, 2009; Shenhar et al., 2007).

Successful project management contributes to the implementation of innovative ideas and influences the creative problem-solving process at much earlier stages of project development (Caughron and Mumford, 2008); enables accurate cost estimates to be produced; acts as an early warning system and keeps the project team focused (Gelbard and Carmeli, 2009); and reduces risks and the time required to complete the project (Söderholm, 2008). Successful management can help in the development of strategic information for customers to address risk and decide whether to commit resources to maximize the likelihood of a successful project (Griffith, Gibson, Hamilton, Tortora, and Wilson, 1999).

Project management plays a crucial role in the initial steps of the project. During the planning phase, the project managers and their teams meet, except when the project is virtual, to effectively plan their execution of the project. The activities entail planning the scope, cost, schedule, risks, quality, communication, human resources, contract, and procurement (PMI, 2015). Planning these aforementioned knowledge area perspectives requires the completion of a work breakdown structure to define the work necessary to produce the deliverables (Aramo-Immonen and Vanharanta, 2009).

In conclusion, it is expected project management to have an effect on success, this hypothesis is accepted and the relations between these two CSFs are statistically validated. Previous studies found in the literature also support the existence of this relation.

The path coefficient is an important indicator which represents the hypothesized relationships among the constructs. As indicated above, all constructs have a significant positive impact on the constructs which are stated in the hypotheses. Hypothesis 1 stated a positive relationship between Governance and Team. Path coefficient of H1 is 0,461 which is the second highest value on the table and refers to a high positive impact. Hypothesis 2 stated a positive relationship between Governance and Technology. The highest impact seen on the research model was beet these two constructs with 0,592, accordingly, H2 was strongly supported. The path coefficient of H3 was 0,346 which is between Governance and Project Management. This value also indicates a statistically significant positive impact. Project Definition was expected to positively affect Project Management which is stated in H4. H4 was also supported with a path coefficient value of 0,212. H5 stated a positive relationship between Team and Success; this is also supported with a path coefficient value of 0,265. It is expected Technology to affect Success positively and that is stated in H6. The hypothesis is statistically significant with the third highest path coefficient value of 0,420. H7 indicated that there is a positive relationship between Project Management and Success. This is also accepted with 0,233, which refers to a statistically significant relationship.

Discussions

This study aims to provide empirical evidence on the factors contributing to the success of Big Data projects and generate a reliable and valid measurement scale and propose a research model. Hence, this research is significant, since the success rates of Big Data projects is largely subjective, so more substantive results can contribute to the project management body of knowledge.

The main purpose of this research was to build on the current diverse literature around Big Data by contributing discussion and data that allow common agreement on definition, characteristics, and factors that influence successful Big Data projects. The research questions being investigated are based on the argument establishing Big Data be used as a tool for the organization by which to develop and create enterprise-wide efficiencies. The researcher tries to explore the following question in this research, "What are the critical success factors that impact perceived project success of Big Data projects?" (qualitative research question) and "What are the relationships among the critical success factors?" (quantitative research question). In this context, CSFs and several hypotheses mentioning relations among the CSF are examined. The relations between CSFs are visualized and tested in a relational model via SEM.

This dissertation puts forward a grounded contribution in terms of CSFs, which is, to our knowledge, the first study in the field of Big data research. Through detailed thematic analysis, concepts and categories were developed, leading the researcher to identify 25 CSF statements representing the CSF categories of Big Data projects. Actually, entire research was based on these qualitative results. The reliable Big Data Success Scale and validated model are developed in the light of 25 CSF and 6 Success statements.

After a series of analysis processes, five CSFs (5 group of CSF categories) have emerged within the research framework. These are Governance, Project Definition, Project Management, and Team. The mean scores of the constructs revealed that Big Data professionals assess Team as the most critical success factor. Governance has the second highest mean score. Governance is followed by Technology, Project Definition, and Project management respectively.

According to item statistics, TM5 (It is critical that the project team is trained or educated on Big Data.) was the highest rated item by Big Data professionals. That means the participants most strongly agreed on this item. This indicated that they see Big Data related education as the most critical requirement for successful completion of a Big Data project. Education is one of the most important issues for niche IT projects. Big Data projects are specially featured projects which involves expertness and know-how. Accordingly, it is expected an education based item to rate the highest mean (Agreement) score. Second highest rated item is TM4 (It is critical that people in the project team have the necessary technical skills.) with 6,421, followed by TM6 (It is critical that people in the team communicate with each other in a healthy way.), TM7 (Healthy interaction between project team end users is critical.) and TM3 (It is critical that people in the project team have analytical thinking skills.) respectively. The findings reveal crucial issues from the point of Big Data professionals' view. The team should consist of people who have required skills set and education for Big Data projects. Skills are followed by communication-related issues. The interaction between the teammates and team-end user communication is very important to reach success.

The key finding of this research is: when taken as a whole, the CSFs of Big Data projects emerging within this study, Governance, Team, Technology, Project Management and Project Definition have positive and significant effects on project success. This finding is consistent with results found in the body of literature for other domains. Path coefficients demonstrate that Technology is the most effective CSF on Success. Also, the strongest effect on the research model is between Governance and Technology. Accordingly, Governance and Technology relationship is an important predictor of Big Data project success. Governance's effect on Team is also very high and Team's impact on Success is the second highest path coefficient values after effect of Governance on Technology. These results indicate that Governance's role is determined by Success in terms of Technology and Team issues.

The second important contribution of this study is the development of a reliable and valid scale for empirically measuring Big Data project success.

Table 52: Scale Quality Criterion

	Cronbach's Alpha	**Composite Reliability**
Governance	0,786	0,862
Project Definition	0,813	0,864
Project Management	0,837	0,874
Success	0,896	0,921

Team	0,856	0,890
Technology	0,876	0,904

Due to the limitations of Cronbach's alpha, composite reliability (also known as Dhillon-Goldstein Rho) is more appropriate to be used for PLS-SEM as it takes into account the different outer loadings of the indicator variables. The composite reliability refers to how well of a construct is explained by its own indicators. This measure provides a value, which ranges between 0 and 1. Composite Reliability and Cronbach's alpha measures of 0,7 and above indicates sufficient convergence or internal consistency (Nunnally, 1978; Gefen, Straub and Boudreau, 2000). The results, presented in Table 52, indicate that the developed scale meets the quality criterion in respect to limits found in the literature. This scale can be re-used and the results can be compared in further studies.

Lastly, Big Data Success Model is tested via PLS-SEM method and statistically validated after a series of analyses.

Table 53: Model Quality Criteria

	R Square
Project Management	0,206
Success	0,508
Team	0,212
Technology	0,351

Table 53 summarizes the R-square scores for the independent variables, shown in the research model. R-square value for Technology is 0,351, which indicates 35% of the variance for Technology is driven by Management factor. On the other hand, the R-square value for Success is 0,518, which indicates 51,8% of success is accounted by Team, Technology, Project Management, and Project Definition. The impact of management on team is not that large with the R-square value of 0,212, accounting 21% of the variance from management. Besides, the R-square value for project management is only 0,20, accounting 20% of the variance from management and project definition.

The results indicate that the research model explained %50,8 of Big Data project success with the given constructs (CSFs). This is a very high explanation rate according to methodologically similar studies and well-known models found in the literature (Fishbein and Ajzen, 1975; Davis, Begozzi and Warshaw, 1992; Taylor and Todd, 1995a; 1995b; Rogers, 1995, Bandura, 1986; DeLone and McLean, 1992).

Implications

This study can be beneficial for both academic and practical perspectives. In terms of academic contribution, original research goal is to close the gap, found in the literature, regarding Big Data project success. Relational representation of critical success factors in a statistical model is a new approach for both Big Data and critical success factors literature. The methodology of this study strengthens the findings. Several semi-structured interviews and a two-round Delphi study are conducted to enlighten and reveal the critical success factors of Big Data implementations. The predicted relationships among these factors are visualized in the research model and subsequently, quantitative data is also gathered from 827 big data professionals to validate the theoretical research model statistically.

The results could extend the IS success model by introducing critical success constructs in IS implementations. While the IS success model (DeLone and McLean, 1992, 2003) includes concepts of information and data quality, service quality, and system quality, its weakness lies that it neglects organizational factors such as management support, team or project related issues.

There are also practical contributions for Big Data project owners and top management teams. The practical usage of this study can help organizations to identify factors contributing to the success or failure of Big Data projects. The research is based on the knowledge and experience of a great group of Big Data experts and workers. Statistically significant results and validated relations among the CSFs promises to take smarter steps while planning a Big Data project.

When organizations plan to start Big Data projects, it is crucial to understand the critical factors chain of the project process and bear in mind that the ultimate outcome of the critical factors chain is the successful completion of the Big Data project. Project success is the corresponding measure for the results of the critical factors chain. According to a report from Gartner (2017), 60% of the Big Data projects end with disappointment. Some experts claim that reality is not good and 85% of the Big Data projects do actually fail (techrepublic, 2017). This picture gives us an opinion about the challenges the industry faces in order to reach success.

The study could contribute to professional practice by assisting top management teams with identifying possible problem areas when implementing Big Data projects. In addition, this study could provide Big Data professionals with an increased understanding of how Big Data projects are impacted by the presence or absence of suggested CSFs. Big Data professionals may also be prepared to explain how CSFs are necessary for successful project completion.

The study is part of the broader field of Big Data and business intelligence initiatives, where organizations use these technologies as part of their data and information management strategy in order to achieve enhanced decision-making capabilities.

The study contributed to the body of literature by describing which organizational factors are related to Big Data project success and investigated the relationship among them. Thus, this research uncovers the CSFs of Big Data projects, which we hope to help Big Data project owners to create a better project plan with better chances to meet these expectations.

Conclusion

The current research data supported associations among 5 critical success factors and Big Data project success. In the current research, Technology was supported as being most associated with Big Data. Technology is followed by the construct of the Team. In the light of these findings, enterprises can allocate their resources in a clever and effective way for successful completion of their Big Data project. This ends up with a thrifty management with less investment and more profit.

The current exploratory mixed methods study used the validated Big Data Success Scale to establish the association strength and direction among 5 critical success factors and project success related to organizations located in Turkey. The study also established the effects on relationships via structural equation modeling. 912 respondents participated in the research and analyses are conducted with 827 refined, clear data. The conclusions established an association among 5 critical success factors, which are Governance, Team, Technology, Project Management, and Project Definition. Organizational leaders might use these conclusions to predict project success and promote project management best practices within Big Data team members. Organizational leaders might use the resulting relationships, to predict project success based on the organization’s current resources.

The results provide original CSF dimensions and several sub-dimensions for researchers and practitioners. The research also put forward the common perception that Big Data projects would require technological advances and infrastructure and an education-driven project team to thrive and be successful as a false claim. From a practical and professional aspect, the project managers need not to be concerned that the organization should only be a traditional data environment. Also seen in the demographics of the type of industry, Big Data projects are being carried out in an array of different types of organizations like Communications, Media and Entertainment, Education, Energy, etc.

Theoretical implications are also valuable in terms of expanding Big Data literature with a reliable and valid scale. Researchers can benefit from the scale and use it in order to analyze the trend in their country and this can contribute to international comparisons studies.

Assumptions and Limitations

This study included several assumptions regarding data gathering and analysis. We assume that the participants answered honestly our questions in the semi-structured interviews, Delphi study and CATI survey. They didn't seem to have any bias while answering the questions and, we determine that the subjects also carefully read or listened to the questions. They had basic knowledge of the premise of each question as provided as instructions in the question. Since it is a convenience sample, we assume that it is predominantly representative of the total population of Big Data professionals.

This study encompassed the following limitations: The findings could not be necessarily generalizable to the entire population. The survey was completely anonymous and conducted only in Turkey.

A few points regarding implications for the study to keep in mind are: (a) there were only a small number of experts, who were attended the Delphi study and they were all found via professional and personal contacts of the researcher, (b) the Delphi categorizing and inspection of themes was conducted solely by the researcher and subject to interpretation, (c) the survey questions were formed by the researcher based mostly from literature review and expert opinions, (d) the survey questions had multiple questions, measuring similar characteristics and that may have distributed the impact of some of the factors and (e) anonymity was a very important factor to the experts. Many didn't want to attend the interviews or Delphi nor did they want to be recorded. The researcher provided as much leeway as possible in answering questions, opting out of the study and minimizing the use of competitive knowledge.

Future Work

This exploratory study is among the preliminary research efforts toward critical success factors in Big Data projects from the perspective of people working on such projects in the industry including Project Managers, Team Members, Customers, Organization management, etc. It provides a starting point for future research related to success factors in Big Data projects as perceived by industry professionals.

Indeed, every Big Data project is unique with the methodology followed. Every project has its own unique features. The impact of these critical success factors on every stage of project lifecycle needs to be justified. The results might include factors that affect differently at different stages. Some factors may not even apply at every stage of project execution.

Big Data projects are carried out worldwide. Due to this global nature, it is necessary to gather data about Big Data projects being executed in different countries. The analysis should be carried out if the method of executing these projects differs from country to country or differ from industry to industry. New success factors can be discovered that can provide insight into the success of Big Data projects in different markets and cultures.

Future studies can also use a qualitative or mixed-method approach. Such an approach can help to understand and carry out research deeper into other aspects of Big Data project implementation. Research can include the impact of technology being used on the method of project execution. It is important to understand if the use of specific technologies or algorithms like can change the way the Big Data project is being executed. In this case, a newly revised list of critical success factors can be found.

Also, the impact of these factors, while implementing six sigma methodologies for the Big Data projects, can be studied. Since lean six sigma deals with reducing waste from a process, it can be combined with these critical success factors that help to identify only the processes that are of greater importance for the project, thus removing waste processes. Also, based on the results of the analysis of critical success factors on project success, a dashboard can be built that can help project managers understand the areas that are lacking during the execution of projects. This can help bring forth the factors that are highly contributing towards project success, as well as those who have a scope for improvement. A graphical view of the dashboard will lead to a better and quicker understanding for professionals working on the project ranging from Project Team, Organization Management, to even the Customer. The research on CSF in Big Data projects has identified different CSF categorized mostly in groups of People, Technology, Process, and Organization, but not categorized these factors in an accepted framework such as PMBOK (Project Management Book of Knowledge) or CMMI (Capability Maturity Model Integration). Also, a study reveals that authors are not quite in agreement in their classification style (Eybers and Hattingh, 2017).

REFERENCES

Books

Allen, M., & Cervo, D. (2015). *Multi-domain master data management: Advanced MDM and Data Governance in Practice*. Morgan Kaufmann.

Angleitner, A., & Wiggins, J. S. (Eds.). (2012). *Personality assessment via questionnaires: Current issues in theory and measurement*. Springer Science & Business Media.

Axson, D. A. (2007). Best Practices in Planning and Performance Management: From Data to Decisions (Wiley Best Practices).

Baker, B. N., Murphy, D. C., & Fisher, D. (1997). Factors affecting project success. *Project management handbook*, 902-919.

Balian, E. S. (1994). *The graduate research guidebook: A practical approach to doctoral/masters research*. Univ Pr of Amer.

Barclay, C. (2015). Critical success factors in knowledge discovery and data mining projects. In K.-M. Osei-Bryson & C. Barclay (Eds.), *Knowledge Discovery Process and Methods*. Boca Raton, FL: CRC Press.

Barclay, C. (2016). Measuring the success of data mining projects: An exploratory application of the project performance scorecard. In C. Barclay & K.-M. Osei-Bryson (Eds.), *Strategic Project Management: Contemporary Issues and Strategies for Developing Economies*. Boca Raton, FL: CRC Press.

Basham, L. M. (2010). *Presidents as Transformational or Transactional Leaders in Higher Education*. ProQuest LLC. 789 East Eisenhower Parkway, PO Box 1346, Ann Arbor, MI 48106.

Charmaz, K. (2006). *Constructing grounded theory: A practical guide through qualitative analysis*. Sage.

Cleland, D. I., & King, W. R. (1975). *Systems analysis and project management*. McGraw-Hill.

Clements, J. P. (2008). *Effective project management*. Evans Publishing Group.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences 2nd edn.

Cooper, D. R., & Schindler, P. S. (2003). Business Research Methods (8 th edn.) McGrawHill: New York.

Corbin, J., & Strauss, A. 2015. *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*.

Creswell, J. W. (2009). Mapping the field of mixed methods research.

Creswell, J. W., & Poth, C. N. (2017). *Qualitative inquiry and research design: Choosing among five approaches* (4th ed.). Sage publications.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Holt, Rinehart and Winston, 6277 Sea Harbor Drive, Orlando, FL 32887.

Dalkey, N.C.; Rourke, D.L.; Lewis, R.; Snyder, D. (1972). *Studies in the quality of life*. Lexington, Massachusetts: Lexington Books.

Daniel, J. (2012). *Sampling essentials: Practical guidelines for making sampling choices*. Thousand Oaks, CA: Sage Publications, Inc.

Davenport, T. H., Barth, P., & Bean, R. (2012). *How'big data'is different*. MIT Sloan Management Review.

Delen, D. (2015). *Real-world data mining: Applied business analytics and decision making*. Upper Saddle River, NJ: Pearson Education.

Demchenko, A. P. (2013). *Ultraviolet spectroscopy of proteins*. Springer Science & Business Media.

Dinsmore, P. C., & Cabanis-Brewin, J. (Eds.). (2006). *The AMA handbook of project management*. Amacom Books.

Drucker, P. (1954). The principles of management. *New York*.

Erikson, E. H. E. (1950). *Childhood and society*. New York: Norton.

Evers, N., Cunningham, J., & Hoholm, T. (2014). *Technology entrepreneurship: bringing innovation to the marketplace*. Macmillan International Higher Education.

Galbraith, J. R. (2014). *Designing organizations*. Jossey-Bass & Pfeiffer Imprints, Wiley & Sons

Gallegos, F., Senft, S., Manson, D. P., & Gonzales, C. (2004). *Information technology control and audit* (2nd ed.). Boca Raton, FL: Auerbach. Garson,

Garrett, G. (2007). *World class contracting.* Riverwoods, IL: CCH-Wolters Kluwer Business.

Gido, J., & Clements, J. (2006). *Successful project management.* Mason, OH: Thomson South-Western.

Greene, J. C. (2007). *Mixed methods in social inquiry* (Vol. 9). John Wiley & Sons.

Hair, Joe F., Black, William C., Babin, Barry J., Anderson, Rolph E., and Ronald L. Tatham (2010), *Multivariate Data Analysis*, Vol.7.Upper Saddle River, Prentice Hall.

Hair Jr, J. F., Hult, G. T. M., Ringle, C., & Sarstedt, M. (2016). *A primer on partial least squares structural equation modeling (PLS-SEM)*. Sage Publications.

Head, Thomas C. (2009). Strategic Organization Development: A Failure of True Organization Development? In T. Yaeger & P. Sorensen (Eds.), *Strategic Organization Development: Managing Change for Success.* (pp 23-42). Champaign: Information Age Publishing.

Hedeman, B., Heemst, G., & Fredriksz, H. (2005). *Best practice: Project management based on prince2.* San Antonio, TX: Van Haren.

Henry, J., & Joel, H. (2004). *Software project management: A real-world guide to success*. Pearson/Addison Wesley.

Henseler, J., Ringle, C. M., & Sinkovics, R. R. (2009). The use of partial least squares path modeling in international marketing. In *New challenges to international marketing* (pp. 277-319). Emerald Group Publishing Limited.

Johnson, B., & Christensen, L. (2012). *Educational research: Quantitative, qualitative, and mixed approaches*. Sage.

Kaplan, D. (2008). *Structural equation modeling: Foundations and extensions* (Vol. 10). Sage Publications.

Kendrick, T. (2003). *Identifying and managing project risk: Essential tools for failure-proofing your project.* New York: AMACOM.

Kerzner, H. (2004). *Advanced project management: Best practices on implementation*. John Wiley & Sons.

Kerzner, H. (2006). Project Management: A Systems Approach to Planning, Scheduling and Controlling. *Editorial John Wiley, Hoboken, New Jersey, ISBN, 471225770*.

Kerzner, H. R. (2014). *Lessons learned from project disasters: Case studies and techniques for overcoming project failure*. Somerset, NJ: John Wiley & Sons.

Kline, R. B. (2005). Methodology in the social sciences.

Larose, D. T., & Larose, C. D. (2014). *Discovering knowledge in data: an introduction to data mining*. John Wiley & Sons.

Laudon, K. C., & Laudon, J. P. (2016). *Management information system*. Pearson Education India.

Leach, L. P. (2005). *Lean project management: eight principles for success*. Advanced Projects, Incorporated.

Leavitt, H. J. (1965). Applied Organizational Change in Industry, Structural, Technological and Humanistic Approaches. *Handbook of organizations, 264*.

Lewis, J. P. (2007). *Fundamentals of project management* (3rd ed.). New York: AMACOM.

Linoff, G. S., & Berry, M. J. (2011). *Data mining techniques: for marketing, sales, and customer relationship management*. John Wiley & Sons.

Linstone, H. A., & Turoff, M. (Eds.). (1975). *The delphi method*(pp. 3-12). Reading, MA: Addison-Wesley.

Lipsey, M. W. (1990). *Design sensitivity: Statistical power for experimental research* (Vol. 19). Sage.

Lock, D. (1984), *Project Management*, St Martins Press, New York, NY.

Lohmöller, J. B. (1989). Predictive vs. structural modeling: Pls vs. ml. In *Latent Variable Path Modeling with Partial Least Squares* (pp. 199-226). Physica, Heidelberg.

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

Lucas Jr, H. C. (1999). *Information technology and the productivity paradox: Assessing the value of investing in IT*. Oxford University Press.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011). *Big data: The next frontier for innovation, competition, and productivity*. McKinsey Global Institute.

Martin, C. C. 1976. *Project management*, New York: Amaco.

Marz, N., & Warren, J. (2015). *Big Data: Principles and best practices of scalable real-time data systems*. New York; Manning Publications Co..

Mayer-Schönberger, V., & Cukier, K. (2014). *Learning with big data: The future of education*. Houghton Mifflin Harcourt.

Merriam, S. B. (1988). *Case study research in education: A qualitative approach*. Jossey-Bass.

Mertler, C. A., & Vannatta, R. A. (2005). *Advanced and multivariate statistical methods: Practical application and interpretation* (3rd ed.). Los Angeles, CA: Pyrczak.

Miles, D. (2003).*The 30-second encyclopedia of learning and performance: A trainer's guide to theory, terminology, and practice*. New York, NY: American Management Association.

Morris, P., & Pinto, J. K. (Eds.). (2007). *The Wiley guide to project, program, and portfolio management* (Vol. 3). John Wiley & Sons.

Narayanan, A., Huey, J., & Felten, E. W. (2016). A precautionary approach to big data privacy. In *Data protection on the move* (pp. 357-385). Springer, Dordrecht.

Nunnally, J. (1978). Psychometric methods.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric Theory (McGraw-Hill Series in Psychology)* (Vol. 3). New York: McGraw-Hill.

O'Dell, C. S., O'dell, C., Grayson, C. J., & Essaides, N. (1998). *If only we knew what we know: The transfer of internal knowledge and best practice*. Simon and Schuster.

Pallant, J. (2007). SPSS survival manual, 3rd. *Edition. McGrath Hill*.

Parasuraman, A., Grewal, D., & Krishnan, R. (2006). *Marketing research*. Cengage Learning.

Parchoma, G. (2007). *Faculty integration of computer-mediated learning technologies into teaching praxis* (pp. 1-256). The University of Saskatchewan.

Patton, M. Q. (1990). *Qualitative evaluation and research methods*. SAGE Publications, inc.

Phillips, J. (2004). *IT project management: on track from start to finish*. McGraw-Hill, Inc..

Privitera, G. J. (2015). *Research methods for the behavioral sciences*. Sage Publications.

Rijmenam, M. (2014). *Think bigger: Developing a successful big data strategy for your business*. Amacom.

Rogers, E. M. (1995). Diffusion of Innovations: modifications of a model for telecommunications. In *Die diffusion von innovationen in der telekommunikation* (pp. 25-38). Springer, Berlin, Heidelberg.

Roldán, J. L., & Sánchez-Franco, M. J. (2012). Variance-based structural equation modeling: guidelines for using partial least squares in information systems research. In *Research methodologies, innovations and philosophies in software systems engineering and information systems* (pp. 193-221). IGI Global.

Santos, J. R. A. (1999). Cronbach's alpha: A tool for assessing the reliability of scales. *Journal of extension*, *37*(2), 1-5.

Sas Institute. (1990). *SAS/STAT user's guide: version 6* (Vol. 2). Sas Inst.

Sekaran, U. (2003). *Research methods for business* . Hoboken.

Shenhar, A. J., & Dvir, D. (2007). *Reinventing project management: the diamond approach to successful growth and innovation*. Harvard Business Review Press.

Smith, J. A. (Ed.). (2015). *Qualitative psychology: A practical guide to research methods*. Sage.

Spector, P. E. (1992). *Summated rating scale construction: An introduction* (No. 82). Sage.

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics*. Allyn & Bacon/Pearson Education.

Tabachnick, B. G., Fidell, L. S., & Osterlind, S. J. (2001). Using mulfivariate stafisfics. *Needham Heights,* MA: Allyn Bacon.

Thomas, J. R., Silverman, S., & Nelson, J. (2015). *Research methods in physical activity, 7E*. Human kinetics.

Vatsalan, D., Sehili, Z., Christen, P., & Rahm, E. (2017). Privacy-preserving record linkage for big data: Current approaches and research challenges. In *Handbook of Big Data Technologies* (pp. 851-895). Springer, Cham.

Vinod, B. (2013). Leveraging BIG DATA for competitive advantage in travel. *Journal of revenue and pricing management*, *12*(1), 96-100.

Weisbord, M. R. (2012). *Productive workplaces: Dignity, meaning, and community in the 21st century*. John Wiley & Sons.

White, T. (2012). *Hadoop: The definitive guide*. " O'Reilly Media, Inc.".

Williams, S., & Williams, N. (2010). *The profit impact of business intelligence*. Elsevier.

Wysocki, A. (2007). *Writing new media: Theory and applications for expanding the teaching of composition*. University Press of Colorado.

Wysocki, R. K. (2004). *Project Management Process Improvement*, Artech House. PMI.

Yaremko, Robert M.(1986), ed. *Handbook of research and quantitative methods in psychology: For students and professionals.* Psychology Press.

Yin, R. K. (2003). Case study research: Design and methods (3rd ed., Vol. 5). Thousand Oaks, CA: Sage.

Yin, R. K. (2011). *Applications of case study research*. Sage.

Articles

Abbasi, A., Sarker, S., & Chiang, R. H. (2016). Big data research in information systems: Toward an inclusive research agenda. *Journal of the Association for Information Systems*, *17*(2).

Abdekhoda, M., Ahmadi, M., Gohari, M., & Noruzi, A. (2015). The effects of organizational contextual factors on physicians' attitude toward adoption of Electronic Medical Records. *Journal of Biomedical Informatics*, 53, 174-179. doi: 10.1016/j.jbi.2014.10.008

Acker, O., Gröne, F., Blockus, A., & Bange, C. (2011). In-memory analytics–strategies for real-time CRM. *Journal of Database Marketing & Customer Strategy Management*, *18*(2), 129-136.

Adam, S. (2009). The four stages of effective team-building. *Journal of Training & Management Development Methods*, 23(1), 317-320.

Agarwal, R., & Dhar, V. (2014). Big data, data science, and analytics: The opportunity and challenge for IS research.

Ahangama, S., & Poo, C. C. D. (2015). Improving health analytic process through project, communication and knowledge management.

Ahimbisibwe, A., Cavana, R. Y., & Daellenbach, U. (2015). A contingency fit model of critical success factors for software development projects: A comparison of agile and traditional plan-based methodologies. *Journal of Enterprise Information Management*, *28*(1), 7-33.

Ahmad, J., Muhammad, K., Lloret, J., & Baik, S. W. (2018). Efficient Conversion of Deep Features to Compact Binary Codes using Fourier Decomposition for Multimedia Big Data. *IEEE Transactions on Industrial Informatics*.

Ahmad, N., Haleem, A., & Syed, A. A. (2012). Compilation of critical success factors in implementation of enterprise systems: a study on Indian organisations. *Global journal of flexible systems management*, *13*(4), 217-232.

Ajila, S. A., & Wu, D. (2007). Empirical study of the effects of open source adoption on software development economics. *Journal of Systems and Software*, *80*(9), 1517-1529.

Ajzen, I. (1985). From intentions to actions: A theory of planned behavior. In *Action control* (pp. 11-39). Springer, Berlin, Heidelberg.

Ajzen, I. (1991). The theory of planned behavior. *Organizational behavior and human decision processes*, *50*(2), 179-211.

Akkaya, G. C., & Uzar, C. (2011). Data mining: Concept, techniques, and applications. *GSTF Business Review (GBR)*, 1(2), 47-50. doi: 10.5716_2010-4804_1.2.50

Akkermans, H., & van Helden, K. (2002). Vicious and virtuous cycles in ERP implementation: a case study of interrelations between critical success factors. *European journal of information systems*, *11*(1), 35-46.

Akter, S., & Wamba, S. F. (2016). Big data analytics in E-commerce: a systematic review and agenda for future research. *Electronic Markets*, *26*(2), 173-194.

Al-Ahmad, W., Al-Fagih, K., Khanfar, K., Alsamara, K., Abuleil, S., & Abu-Salem, H. (2009). A Taxonomy of an IT Project Failure: Root Causes. *International Management Review*, 5(1), 93-104,106. doi: 1728780921

Al Nuaimi, E., Al Neyadi, H., Mohamed, N., & Al-Jaroodi, J. (2015). Applications of big data to smart cities. *Journal of Internet Services and Applications*, *6*(1), 25.

Alexander, F. J., Hoisie, A., & Szalay, A. (2011). Big data. *Computing in Science & Engineering*, *13*(6), 10-13.

Alexandrov, A. V., Pullicino, P. M., Meslin, E. M., & Norris, J. W. (1996). Agreement on disease-specific criteria for do-not-resuscitate orders in acute stroke. *Stroke*, *27*(2), 232-237.

Allen, B., Bresnahan, J., Childers, L., Foster, I., Kandaswamy, G., Kettimuthu, R., ... & Tuecke, S. (2012). Software as a service for data scientists. *Communications of the ACM*, *55*(2), 81-88.

Almajed, A. I., & Mayhew, P. (2014). An empirical investigation of IT project success in developing countries. In *Science and Information Conference (SAI), 2014* (pp. 984-990). IEEE.

Al-Mashari, M., & Zairi, M. (2000). Revisiting BPR: a holistic review of practice and development. *Business process management journal*, *6*(1), 10-42.

Al-Qirim, N., Tarhini, A., & Rouibah, K. (2017). Determinants of Big Data Adoption and Success. In *Proceedings of the International Conference on Algorithms, Computing and Systems* (pp. 88-92). ACM.

Altuwaijri, M. M., & Khorsheed, M. S. (2012). InnoDiff: A project-based model for successful IT innovation diffusion. *International Journal of Project Management*, 30(1), 37-47. doi: 10.1016/j.ijproman.2011.04.007

Amit, R., & Schoemaker, P. J. (1993). Strategic assets and organizational rent. *Strategic management journal*, *14*(1), 33-46.

Amoako-Gyampah, K. (2004). ERP implementation factors: A comparison of managerial and end-user perspectives. *Business Process Management Journal*, *10*(2), 171-183.

Anderson, E. S., Birchall, D., Jessen, S. A., & Money, A. H. (2006). Exploring project success. *Baltic Journal of Management*, 1, 127-147. Retrieved May 15, 2008, from Emerald database.

Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological bulletin*, *103*(3), 411.

Anderson, S., & Blanke, T. (2012). Taking the long view: from e-science humanities to humanities digital ecosystems. *Historical Social Research/Historische Sozialforschung*, 147-164.

Ang, J., & Teo, T. S. (2000). Management issues in data warehousing: insights from the Housing and Development Board. *Decision Support Systems*, *29*(1), 11-20.

Angelides, D. C. (1999). Project management and good technical and business practices. *Journal of Management in Engineering*, *15*(3), 78-88.

Anggadini, S. D. (2015). The Effect of Top Management Support and Internal Control of the Accounting Information Systems Quality and Its Implications on the Accounting Information Quality. *Information Management and Business Review*, 7(3), 93-102.

Anie, K. A., Jones, P. W., Hilton, S. R., & Anderson, H. R. (1996). A computer-assisted telephone interview technique for assessment of asthma morbidity and drug use in adult asthma. *Journal of clinical epidemiology*, *49*(6), 653-656.

Antonacopoulou, E. P. (2001). The paradoxical nature of the relationship between training and learning. *Journal of Management Studies*, *38*(3), 327-350.

Aramo-Immonen, H., & Vanharanta, H. (2009). Project management: The task of holistic systems thinking. *Human Factors and Ergonomics in Manufacturing & Service Industries*, *19*(6), 582-600.

Arghode, V. (2012). Qualitative and Quantitative Research: Paradigmatic Differences. *Global Education Journal*, *2012*(4).

Argyrous, G. (2005). Statistics for social health and health research: with a guide to SPSS.

Arnott, D. (2008). Success factors for data warehouse and business intelligence systems. *ACIS 2008 Proceedings*, 16.

Ashby, N., Fryirs, K., & Howitt, R. (2015). Prospects for, and Challenges of, Research Design and Training in Cross-Disciplinary Environmental Management Research. *Geographical Research*, *53*(1), 81-94.

Audzeyeva, A., & Hudson, R. (2016). How to get the most from a business intelligence application during the post implementation phase? Deep structure transformation at a UK retail bank. *European Journal of Information Systems*, *25*(1), 29-46.

Ayankoya, K., Calitz, A., & Greyling, J. (2014, September). Intrinsic relations between data science, big data, business analytics and datafication. In *Proceedings of the Southern African Institute for Computer Scientist and Information Technologists*

Annual Conference 2014 on SAICSIT 2014 Empowered by Technology (p. 192). ACM.

Baccarini, D., & Collins, A. (2003). Critical success factors for projects. In *Proceedings of the 17th ANZAM Conference*.

Bagchi, S., Kanungo, S., & Dasgupta, S. (2003). Modeling use of enterprise resource planning systems: a path analytic study. *European Journal of Information Systems*, *12*(2), 142-158.

Bagozzi, R. P., Davis, F. D., & Warshaw, P. R. (1992). Development and test of a theory of technological learning and usage. *Human relations*, *45*(7), 659-686.

Bagozzi, R. P., & Yi, Y. (1989). The degree of intention formation as a moderator of the attitude-behavior relationship. *Social Psychology Quarterly*, 266-279.

Bagozzi, R. P., Yi, Y., & Phillips, L. W. (1991). Assessing construct validity in organizational research. *Administrative science quarterly*, 421-458.

Bailey, J. E., & Pearson, S. W. (1983). Development of a tool for measuring and analyzing computer user satisfaction. *Management science*, *29*(5), 530-545.

Bandura, A. (1989). Regulation of cognitive processes through perceived self-efficacy. *Developmental psychology*, *25*(5), 729.

Baray, S., Hameed, S., & Badii, A. (2006). Analyzing the Effectiveness of Implementing Enterprise Resource Planning in the Printing Industry. In *European and Mediterranean Conference on Information Systems (EMCIS)* (pp. 6-7).

Baray, S., Hameed, S., & Badii, A. (2008). Analysing the factors responsible for effectiveness of implementation and integration of enterprise resource planning systems in the printing industry. *Journal of enterprise information management*, *21*(2), 139-161.

Barrett, P. (2007). Structural equation modelling: Adjudging model fit. *Personality and Individual differences*, *42*(5), 815-824.

Bartlett, M. S. (1954). A note on the multiplying factors for various χ 2 approximations. *Journal of the Royal Statistical Society. Series B (Methodological)*, 296-298.

Bartlett, J. W., Kotrlik, C. C., & Higgins. (2001). Organizational Research: Determining Appropriate Sample Size in Survey Research. *Information Technology, Learning, and Performance Journal*, *19*(1).

Bazeley, P. (2009). Integrating data analyses in mixed methods research.

Bean, R. A. N. D. Y., & Kiron, D. (2013). Organizational alignment is key to big data success. *MIT Sloan Management Review*, *54*(3), 1-6.

Bearden, W. O., Hardesty, D. M., & Rose, R. L. (2001). Consumer self-confidence: Refinements in conceptualization and measurement. *Journal of Consumer Research*, *28*(1), 121-134.

Beath, C., Becerra-Fernandez, I., Ross, J., & Short, J. (2012). Finding value in the information explosion. *MIT Sloan Management Review*, *53*(4), 18.

Becker, G. E., & Roberts, T. (2009). Do we agree? Using a Delphi technique to develop consensus on skills of hand expression. *Journal of human lactation*, *25*(2), 220-225.

Becker, J. M., Klein, K., & Wetzels, M. (2012). Hierarchical latent variable models in PLS-SEM: guidelines for using reflective-formative type models. *Long Range Planning*, *45*(5-6), 359-394.

Besner, C., & Hobbs, B. (2008). Project management practice, generic or contextual: A reality check. *Project Management Journal*, *39*(1), 16-33.

Beyer, M. A., & Laney, D. (2012). The importance of ‘big data’: a definition. *Stamford, CT: Gartner*, 2014-2018.

Bharadwaj, A., El Sawy, O., Pavlou, P., & Venkatraman, N. (2013). Digital business strategy: toward a next generation of insights.

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.

Bhuasiri, W., Xaymoungkhoun, O., Zo, H., Rho, J. J., & Ciganek, A. P. (2012). Critical success factors for e-learning in developing countries: A comparative analysis between ICT experts and faculty. *Computers & Education*, *58*(2), 843-855.

Biernacki, P., & Waldorf, D. (1981). Snowball sampling: Problems and techniques of chain referral sampling. *Sociological methods & research*, *10*(2), 141-163.

Bizer, C., Boncz, P., Brodie, M. L., & Erling, O. (2012). The meaningful use of big data: four perspectives--four challenges. *Acm Sigmod Record*, *40*(4), 56-60.

Boehm, B., & Turner, R. (2005). Management challenges to implementing agile processes in traditional development organizations. *IEEE software*, *22*(5), 30-39.

Boja, C., Pocovnicu, A., & Batagan, L. (2012). Distributed Parallel Architecture for" Big Data". *Informatica Economica*, *16*(2), 116.

Bole, U., Popovič, A., Žabkar, J., Papa, G., & Jaklič, J. (2015). A case analysis of embryonic data mining success. *International Journal of Information Management*, 35(2), 253-259. doi: 10.1016/j.ijinfomgt.2014.12.001

Bollen, K. A. (2011). Evaluating effect, composite, and causal indicators in structural equation models. *Mis Quarterly*, 359-372.

Boone, C. A., Skipper, J. B., & Hazen, B. T. (2017). A framework for investigating the role of big data in service parts management. *Journal of cleaner production*, *153*, 687-691.

Boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, communication & society*, *15*(5), 662-679.

Boyd, J., Ferrante, A., Brown, A., Randall, S., & Semmens, J. (2017). Implementing privacy-preserving record linkage: welcome to the real world. *International Journal of Population Data Science*, *1*(1).

Bradbury, D. (2011). Data mining with LinkedIn. *Computer Fraud & Security*, *2011*(10), 5-8.

Brady, S. R. (2015). Utilizing and adapting the Delphi method for use in qualitative research. *International Journal of Qualitative Methods*, *14*(5), 1609406915621381.

Brinkmann, B. H., Bower, M. R., Stengel, K. A., Worrell, G. A., & Stead, M. (2009). Large-scale electrophysiology: acquisition, compression, encryption, and storage of big data. *Journal of neuroscience methods*, *180*(1), 185-192.

Brooks, K. W. (1979). Delphi technique: Expanding applications. *North Central Association Quarterly*, *53*(3), 377-85.

Browning, T. R., & Ramasesh, R. V. (2015). Reducing unwelcome surprises in project management. *MIT Sloan Management Review*, *56*(3), 53.

Bryant, R., Katz, R. H., & Lazowska, E. D. (2008). Big-data computing: creating revolutionary breakthroughs in commerce, science and society.

Bullen, C. V., & Rockart, J. F. (1981). A primer on critical success factors.

Burg, N. (2014). How Big Data Will Help Save Healthcare. *Forbes Magazine*, *10*.

Brancheau, J. C., Janz, B. D., & Wetherbe, J. C. (1996). Key issues in information systems management: 1994-95 SIM Delphi results. *MIS quarterly*, 225-242.

Brown, B., Chui, M., & Manyika, J. (2011). Are you ready for the era of 'big data'. *McKinsey Quarterly*, *4*(1), 24-35.

Brynjolfsson, E. (1993). The productivity paradox of information technology. *Communications of the ACM*, *36*(12), 66-77.

Brynjolfsson, E., & McAfee, A. (2013). The great decoupling. *New Perspectives Quarterly*, *30*(1), 61-63.

Brynjolfsson, E., & Hitt, L. (1993). *Is information systems spending productive?: new evidence and new results* (pp. 47-64). Cambridge, MA: MIT Sloan School of Management.

Bughin, J., Chui, M., & Manyika, J. (2010). Clouds, big data, and smart assets: Ten tech-enabled business trends to watch. *McKinsey quarterly*, *56*(1), 75-86.

Bughin, J., Livingston, J., & Marwaha, S. (2011). Seizing the potential of 'big data'. *McKinsey Quarterly*, *4*, 103-109.

Burgess, J., & Bruns, A. (2012). Twitter archives and the challenges of" Big Social Data" for media and communication research. *M/C Journal*, *15*(5).

Buriak, P., & Shinn, G. C. (1989). Mission, initiatives, and obstacles to research in agricultural education: A national Delphi using external decision-makers. *Journal of Agricultural Education*, *30*(4), 14-23.

Calisir, F., & Calisir, F. (2004). The relation of interface usability characteristics, perceived usefulness, and perceived ease of use to end-user satisfaction with enterprise resource planning (ERP) systems. *Computers in human behavior*, *20*(4), 505-515.

Callebaut, W. (2012). Scientific perspectivism: A philosopher of science's response to the challenge of big data biology. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, *43*(1), 69-80.

Calvard, T. S. (2016). Big data, organizational learning, and sensemaking: Theorizing interpretive challenges under conditions of dynamic complexity. *Management learning*, *47*(1), 65-82.

Calvert, C., & Carroll, J. (2005). Training for ERP: Does the IS training literature have value?. *ACIS 2005 Proceedings*, 108.

Campbell, J. P. (1982). Some remarks from the outgoing editor.

Cao, G., & Duan, Y. (2014). A path model linking business analytics, data-driven culture, and competitive advantage.

Cao, M., Chychyla, R., & Stewart, T. (2015). Big Data analytics in financial statement Audits. *Accounting Horizons*, *29*(2), 423-429.

Caracelli, V. J., & Greene, J. C. (1997). Crafting mixed-method evaluation designs. *New directions for evaluation*, *1997*(74), 19-32.

Carr, N. G. (2003). IT doesn't matter. *Educause Review*, *38*, 24-38.

Caughron, J. J., & Mumford, M. D. (2008). Project planning: The effects of using formal planning techniques on creative problem-solving. *Creativity and Innovation Management*, *17*(3), 204-215.

Caulfield, T., Rachul, C., & Zarzeczny, A. (2012). The evolution of policy issues in stem cell research: an international survey. *Stem Cell Reviews and Reports*, *8*(4), 1037-1042.

Caughron, J. J., & Mumford, M. D. (2008). Project planning: The effects of using formal planning techniques on creative problem-solving. *Creativity and Innovation Management*, *17*(3), 204-215.

Chae, B. K. (2015). Insights from hashtag# supplychain and Twitter Analytics: Considering Twitter and Twitter data for supply chain practice and research. *International Journal of Production Economics*, *165*, 247-259.

Chae, J., Thom, D., Jang, Y., Kim, S., Ertl, T., & Ebert, D. S. (2014). Public behavior response analysis in disaster events utilizing visual analytics of microblog data. *Computers & Graphics*, *38*, 51-60.

Chang, V., & Wills, G. (2016). A model to compare cloud and non-cloud storage of Big Data. *Future Generation Computer Systems*, *57*, 56-76.

Chen, C., Young, D., & Zhuang, Z. (2012). Externalities of mandatory IFRS adoption: Evidence from cross-border spillover effects of financial information on investment efficiency. *The Accounting Review*, *88*(3), 881-914.

Chen, C. A., & Hsieh, C. W. (2015). Knowledge sharing motivation in the public sector: the role of public service motivation. *International Review of Administrative Sciences*, *81*(4), 812-832.

Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business intelligence and analytics: from big data to big impact. *MIS quarterly*, 1165-1188.

Chen, H. M., Schütz, R., Kazman, R., & Matthes, F. (2016). Amazon in the air: Innovating with big data at Lufthansa. In *System Sciences (HICSS), 2016 49th Hawaii International Conference on* (pp. 5096-5105). IEEE.

Chen, J., Chen, Y., Du, X., Li, C., Lu, J., Zhao, S., & Zhou, X. (2013). Big data challenge: a data management perspective. *Frontiers of Computer Science*, *7*(2), 157-164.

Chen, M., Mao, S., and Liu, Y. (2014). Big Data: A Survey. *Mob. Networks Appl.*, Vol. 19, No. 2, Pp. 171– 209.

Chin, W. W. (1998). The partial least squares approach to structural equation modeling. *Modern methods for business research*, *295*(2), 295-336.

Chin, W. W., & Newsted, P. R. (1999). Structural equation modeling analysis with small samples using partial least squares. *Statistical strategies for small sample research*, *1*(1), 307-341.

Choi, B. C. (2004). Computer assisted telephone interviewing (CATI) for health surveys in public health surveillance: methodological issues and challenges ahead. *Chronic Diseases and Injuries in Canada*, *25*(2), 21.

Church, A., & Dutta, S. (2013). The promise of big data for OD. *OD Practitioner*, *45*(4), 23-31.

Churchill Jr, G. A. (1979). A paradigm for developing better measures of marketing constructs. *Journal of marketing research*, 64-73.

Cicmil, S., & Hodgson, D. (2006). New possibilities for project management theory: A critical engagement. *Project Management Journal*, *37*(3), 111-122.

Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological assessment*, *7*(3), 309.

Clinton, B. D., & Lummus, R. R. (2000). ERP in institutional manufacturing. *Manage. Account. Quart*, *1*, 18-24.

Cohen, J. (1992). A power primer. *Psychological bulletin*, *112*(1), 155.

Cohen, W. M., & Levinthal, D. A. (1994). Fortune favors the prepared firm. *Management science*, *40*(2), 227-251.

Cole, J. B., Newman, S., Foertter, F., Aguilar, I., & Coffey, M. (2012). Breeding and genetics symposium: Really big data: Processing and analysis of very large data sets. *Journal of animal science*, *90*(3), 723-733.

Comrey, A. L., & Lee, H. B. (1992). Interpretation and application of factor analytic results. *Comrey AL, Lee HB. A first course in factor analysis, 2*.

Connelly, D. R., & Canestraro, D. (2007). *Success factors in intergovernmental information technology projects. Paper presented at the Midwestern Political Science Association*, Palmer House Hotel, Chicago.

Conway, J. M., & Lance, C. E. (2010). What reviewers should expect from authors regarding common method bias in organizational research. *Journal of Business and Psychology*, *25*(3), 325-334.

Cooke-Davies, T. (2002). The "real" success factors on projects. *International journal of project management*, *20*(3), 185-190.

Corbin, J. M., & Strauss, A. (1990). Grounded theory research: Procedures, canons, and evaluative criteria. *Qualitative sociology*, *13*(1), 3-21.

Coyne, I. T. (1997). Sampling in qualitative research. Purposeful and theoretical sampling; merging or clear boundaries?. *Journal of advanced nursing*, *26*(3), 623-630.

Creswell, J. W. (2005). Mixed methods designs. *Educational research: Planning, conducting, and evaluating quantitative and qualitative research*, 509-529.

Creswell, J. W., Klassen, A. C., Plano Clark, V. L., & Smith, K. C. (2011). Best practices for mixed methods research in the health sciences. *Bethesda (Maryland): National Institutes of Health, 2013*, 541-545.

Creswell, J. W., & Plano Clark, V. L. (2011). Choosing a mixed methods design. *Designing and conducting mixed methods research*, 53-106.

Cserháti, G., & Szabó, L. (2014). The relationship between success criteria and success factors in organisational event projects. *International Journal of Project Management*, *32*(4), 613-624.

Custer, R. L., Scarcella, J. A., & Stewart, B. R. (1999). The Modified Delphi Technique--A Rotational Modification. *Journal of vocational and technical education*, *15*(2), 50-58.

Cuzzocrea, A., Song, I. Y., & Davis, K. C. (2011, October). Analytics over large-scale multidimensional data: the big data revolution!. In *Proceedings of the ACM 14th international workshop on Data Warehousing and OLAP* (pp. 101-104). ACM.

Dansion, F., & Griffin, J. (2012). Analytics and the Cloud-the Future is here. *Financial Executive*, *28*(9), 97-99.

Dai, C. X., & Wells, W. G. (2004). An exploration of project management office features and their relationship to project performance. *International Journal of Project Management*, *22*(7), 523-532.

Dalkey, N., & Helmer, O. (1963). An experimental application of the Delphi method to the use of experts. *Management science*, *9*(3), 458-467.

Das, M., Cui, R., Campbell, D. R., Agrawal, G., & Ramnath, R. (2015). Towards methods for systematic research on big data. In *Big Data (Big Data), 2015 IEEE International Conference on* (pp. 2072-2081). IEEE.

Davenport, T. H., Barth, P., & Bean, R. (2012). *How'big data'is different*. MIT Sloan Management Review.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS quarterly*, 319-340.

Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1992). Extrinsic and intrinsic motivation to use computers in the workplace 1. *Journal of applied social psychology*, *22*(14), 1111-1132.

Davis, K. (2014). Different stakeholder groups and their perceptions of project success. *International Journal of Project Management*, 32(2), 189-201. doi: 10.1016/j.ijproman.2013.02.006

Dawson, L., & Van Belle, J. P. (2013). Critical success factors for business intelligence in the South African financial services sector. *South African Journal of Information Management*, *15*(1), 1-12.

de Rond, M. (2012). *There is an I in Team: what elite athletes and coaches really know about high performance*. Harvard Business Press.

Deliverables. (2007). PM Network, 21(7), 19.

DeLone, W. H., & McLean, E. R. (1992). Information systems success: The quest for the dependent variable. *Information systems research*, *3*(1), 60-95.

Delone, W. H., & McLean, E. R. (2003). The DeLone and McLean model of information systems success: a ten-year update. *Journal of management information systems*, *19*(4), 9-30.

Demirkan, H., & Delen, D. (2013). Leveraging the capabilities of service-oriented decision support systems: Putting analytics and big data in cloud. *Decision Support Systems*, *55*(1), 412-421.

DeVellis, R. F. (2016). *Scale development: Theory and applications* (Vol. 26). Sage publications.

DeWalt, D. A., Rothrock, N., Yount, S., & Stone, A. A. (2007). Evaluation of item candidates: the PROMIS qualitative item review. *Medical care*, *45*(5 Suppl 1), S12.

De Wit, A. (1988). Measurement of project success. *International journal of project management*, *6*(3), 164-170.

Dezdar, S., & Ainin, S. (2011). Examining ERP implementation success from a project environment perspective. *Business Process Management Journal*, *17*(6), 919-939.

Dhillon, G., & Caldeira, M. (2008). A bumpy road to success (or not): The case of Project Genesis at Nevada DMV. *International Journal of Information Management*, *28*(3), 222-228.

Diamantopoulos, A. (2011). Incorporating formative measures into covariance-based structural equation models. *Mis Quarterly*, 335-358.

Diamantopoulos, A., & Siguaw, J. A. (2006). Formative versus reflective indicators in organizational measure development: A comparison and empirical illustration. *British Journal of Management*, *17*(4), 263-282.

Dillman, D., Smyth, J., & Christian, L. (2014). Chapter 10: Mail questionnaires and implementation. *Internet, Phone, Mail, and Mixed-Mode Surveys. The Tailored Design Method. Hoboken, NJ: John Wiley & Sons, Inc*, 351-397.

Dimopoulos, S., Krintz, C., & Wolski, R. (2016, December). Big data framework interference in restricted private cloud settings. In *Big Data (Big Data), 2016 IEEE International Conference on*(pp. 335-340). IEEE.

DiStefano, C., Zhu, M., & Mindrila, D. (2009). Understanding and using factor scores: Considerations for the applied researcher. *Practical Assessment, Research & Evaluation*, *14*(20), 1-11.

Dowling, G. R. (1993). Developing your company image into a corporate asset. *Long range planning*, *26*(2), 101-109.

Dvir, D. (2005). Transferring projects to their final users: The effect of planning and preparations for commissioning on project success. *International Journal of Project Management*, *23*(4), 257-265.

Doke, E. R., & Swanson, N. E. (1995). Decision variables for selecting prototyping in information systems development: A Delphi study of MIS managers. *Information & Management*, *29*(4), 173-182.

Dong, L. (2008). Exploring the impact of top management support of enterprise systems implementations outcomes: Two cases. *Business Process Management Journal*, *14*(2), 204-218.

Dong, C. K. B. C., Chuah, K. B., & Zhai, L. (2004). A study of critical success factors of information system projects in China. In *Proceedings of th PMI Research Conference, London*.

Dong, L., Neufeld, D., & Higgins, C. (2009). Top management support of enterprise systems implementations. *Journal of Information Technology*, 24(1), 55-80. doi: 10.1057/jit.2008.21

Dos Santos, B., & Sussman, L. (2000). Improving the return on IT investment: the productivity paradox. *International journal of information management*, *20*(6), 429-440.

Doty, D. H., & Glick, W. H. (1998). Common methods bias: does common methods variance really bias results?. *Organizational research methods*, *1*(4), 374-406.

Du, Z. (2013). Inconsistencies in big data: Cognitive Informatics & Cognitive Computing (ICCI* CC), 2013. In *12th IEEE International Conference. IEEE*.

Dumbill, E. (2013). Making sense of big data. *Big Data.* 1(1).

Duplaga, E. A., & Astani, M. (2003). Implementing ERP in manufacturing. *Information Systems Management* , 20 (3), 68-75.

Dutta, D., & Bose, I. (2015). Managing a big data project: the case of ramco cements limited. *International Journal of Production Economics*, *165*, 293-306.

E Silva, L. C., & Costa, A. P. C. S. (2013). Decision model for allocating human resources in information system projects. *International Journal of Project Management*, *31*(1), 100-108.

Ebner, K., Buhnen, T., & Urbach, N. (2014, January). Think big with Big Data: Identifying suitable Big Data strategies in corporate environments. In *System Sciences (HICSS), 2014 47th Hawaii International Conference on* (pp. 3748-3757). IEEE.

Ehrenstein, V., Nielsen, H., Pedersen, A. B., Johnsen, S. P., & Pedersen, L. (2017). Clinical epidemiology in the era of big data: new opportunities, familiar challenges. *Clinical epidemiology*, *9*, 245.

Elbanna, A. (2013). Top management support in multiple-project environments: an in-practice view. *European Journal of Information Systems*, 22(3), 278-294. doi: 10.1057/ejis.2012.16

Elragal, A. (2014). ERP and big data: the inept couple. *Procedia Technology*, *16*, 242-249.

Erwin, R. (2015). Data literacy: Real-world learning through problem-solving with data sets. *American Secondary Education*, 43(2), 18-26

Esposito, C., Ficco, M., Palmieri, F., & Castiglione, A. (2015). A knowledge-based platform for Big Data analytics based on publish/subscribe services and stream processing. *Knowledge-Based Systems*, *79*, 3-17.

Esteves, J., & Pastor, J. (1999). An ERP lifecycle-based research agenda. In *1st International Workshop in Enterprise Management & Resource Planning*.

Evans, G. L. (2013). A novice researcher's first walk through the maze of grounded theory. *Grounded Theory Review*, *12*(1).

Fernandes, L. M., O'Connor, M., & Weaver, V. (2012). Big data, bigger outcomes. *Journal of AHIMA*, *83*(10), 38-43.

Fesenko, T., & Minaev, D. (2014). Costomer focus in the project communications management (on the example of house building). *Eastern-European Journal of Enterprise Technologies*, *5*(3), 4.

Fesenmaier, D. R., Kuflik, T., & Neidhardt, J. (2016, September). Rectour 2016: workshop on recommenders in tourism. In *Proceedings of the 10th ACM Conference on Recommender systems* (pp. 417-418). ACM.

Finch, C. (2009). Keys to effective personnel management. Ceramic Industry, 159(8), 34-35.

Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention and behavior: An introduction to theory and research*.

Fisher, D., DeLine, R., Czerwinski, M., & Drucker, S. (2012). Interactions with big data analytics. *interactions*, *19*(3), 50-59.

Fletcher, A. & Childon, G. P. (2014). Using the Delphi method for qualitative, participatory action research in health leadership. *International Journal of Qualitative Methods*, 13(1), 1-18.

Florentine, S. (2013). Why are so many IT projects failing. *CIO.*

Ford, J. K., MacCallum, R. C., & Tait, M. (1986). The application of exploratory factor analysis in applied psychology: A critical review and analysis. *Personnel psychology*, *39*(2), 291-314.

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of marketing research*, 39-50.

Forte, P. (1994). Data rich, information poor: Data, information and decision support in the NHS. *European Journal of Information Systems*, *3*(2), 148-154.

Fortune, J., & White, D. (2006). Framing of project critical success factors by a systems model. *International journal of project management*, *24*(1), 53-65.

Fourati-Jamoussi, F., & Niamba, C. N. (2016). An evaluation of business intelligence tools: a cluster analysis of users' perceptions. *Journal of Intelligence Studies in Business*, *6*(1).

Françoise, O., Bourgault, M., & Pellerin, R. (2009). ERP implementation through critical success factors' management. *Business Process Management Journal*, *15*(3), 371-394.

Franklin, K. K., & Hart, J. K. (2007). Idea generation and exploration: Benefits and limitations of the policy Delphi research method. *Innovative Higher Education*, *31*(4), 237-246.

Franks, B. (2012). *Taming the big data tidal wave: Finding opportunities in huge data streams with advanced analytics* (Vol. 49). John Wiley & Sons.

Freedman, S., & Katz, L. (2007). Critical success factors for international projects. *PM world today*, *9*(10), 1-8.

Fretty, P. (2007). Disaster Strikes: Effective planning helps any project, but in disaster relief efforts, it can make the difference between life and death, normality and chaos. *PM Network*, *21*(2), 34.

Gaardboe, R., Nyvang, T., & Sandalgaard, N. (2017). Business intelligence success applied to healthcare information systems. *Procedia Computer Science*, *121*, 483-490.

Gamage, P. (2016). New development: Leveraging 'big data'analytics in the public sector. *Public Money & Management*, 36(5), 385-390.

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, *35*(2), 137-144.

Gao, J., Koronios, A., & Selle, S. (2015). Towards a process view on critical success factors in big data analytics projects.

Garmaki, M., Boughzala, I., & Wamba, S. F. (2016, June). The effect of Big Data Analytics Capability on Firm Performance. In *PACIS* (p. 301).

Garrett Jr, R. P., & Neubaum, D. O. (2013). Top management support and Initial strategic assets: A dependency model for internal corporate venture performance. *Journal of Product Innovation Management*, *30*(5), 896-915.

Garg, R. K., & Singh, T. P. (2006). Management of change-a comprehensive review. *Global Journal of Flexible Systems Management*, *7*(1/2), 45-60.

Gefen, D., Rigdon, E. E., & Straub, D. (2011). Editor's comments: an update and extension to SEM guidelines for administrative and social science research. *Mis Quarterly*, iii-xiv.

Gefen, D., Straub, D., & Boudreau, M. C. (2000). Structural equation modeling and regression: Guidelines for research practice. *Communications of the association for information systems*, *4*(1), 7.

Gehrke, J. (2012). Quo vadis, data privacy?. *Annals of the New York Academy of Sciences*, *1260*(1), 45-54.

Geisser, S. (1974). A predictive approach to the random effect model. *Biometrika*, *61*(1), 101-107.

Geisser, S. (1975). The predictive sample reuse method with applications. *Journal of the American statistical Association*, *70*(350), 320-328.

Gelbard, R., & Carmeli, A. (2009). The interactive effect of team dynamics and organizational support on ICT project success. *International Journal of Project Management*, *27*(5), 464-470.

George, G., Haas, M. R., & Pentland, A. (2014). Big data and management.

George, G., Osinga, E. C., Lavie, D., & Scott, B. A. (2016). Big data and data science methods for management research.

Gerbing, D. W., & Anderson, J. C. (1988). An updated paradigm for scale development incorporating unidimensionality and its assessment. *Journal of marketing research*, 186-192.

Ghobadi, S. (2015). What drives knowledge sharing in software development teams: A literature review and classification framework. *Information & Management*, *52*(1), 82-97.

Glaser, B., & Strauss, A. (1967). The discovery ofgrounded theory. *London: Weidenfeld and Nicholson*, *24*(25), 288-304.

Gobble, M. M. (2013). Creating change. *Research-technology management*, *56*(5), 62-66.

Goes, P. B. (2014). Editor's comments: big data and IS research. *Mis Quarterly*, *38*(3), iii-viii.

Gog, I., Giceva, J., Schwarzkopf, M., Vaswani, K., Vytiniotis, D., Ramalingan, G., ... & Isard, M. (2015). Broom: Sweeping out garbage collection from big data systems. *Young*, *4*, 8.

Gold, J. (1998). Telling the story of organizational effectiveness. *Career Development International Journal*, 3(3), 107.

Gomez, L. F., & Heeks, R. (2016). Measuring the Barriers to Big Data for Development. *Development Informatics Working Paper*, *62*.

Gonzales, R., Wareham, J., & Serida, J. (2015). Measuring the Impact of Data Warehouse and Business Intelligence on Enterprise Performance in Peru: A Developing Country. *Journal of Global Information Technology Management*, *18*(3), 162-187.

Goodwin, K. (2011). *Designing for the digital age: How to create human-centered products and services*. John Wiley & Sons.

Gopalakrishnan, K., Yusuf, Y. Y., Musa, A., Abubakar, T., & Ambursa, H. M. (2012). Sustainable supply chain management: A case study of British Aerospace (BAe) Systems. *International Journal of Production Economics*, *140*(1), 193-203.

Gopalasamy, P., Mansor, Z., Selagor, U., & Tambahan, J. T. (2013). An investigation on project management standard practices in IT organization. *International Journal of Computer Engineering Science (IJCES),* 3(1), 1-12. Retrieved

Gracht, H. (2012). Consensus measurement in Delphi studies Re-opinion and implications for future quality assurance. *Technological Forecasting & Social Change*. 79. 1525–1536.

Grant, R. M. (1996). Prospering in dynamically-competitive environments: Organizational capability as knowledge integration. *Organization science*, *7*(4), 375-387.

Gravetter, F. J., & Wallnau, L. B. (2016). *Statistics for the behavioral sciences*. Cengage Learning.

Green, J., Rutherford, S., & Turner, T. (2009). Best practice in using business intelligence to determine research strategy. *Perspectives*, *13*(2), 48-55.

Griffin, R. (2012). Using Big Data to combat enterprise fraud: to combat fraud, more organizations are thinking big--employing new approaches around Big Data to convert the volumes of information available into useful insight and real action. *Financial executive*, *28*(10), 44-48.

Griffith, A. F., Gibson, G. E., Hamilton, M. R., Tortora, A. L., & Wilson, C. T. (1999). Project success index for capital facility construction projects. *Journal of performance of constructed facilities*, *13*(1), 39-45.

Groves, P., Kayyali, B., Knott, D., & Van Kuiken, S. (2013). The 'big data'revolution in healthcare. *McKinsey Quarterly*, *2*(3).

Grublješič, T., & Jaklič, J. (2015). Business intelligence acceptance: The prominence of organizational factors. *Information Systems Management*, *32*(4), 299-315.

Guan, J., & He, Y. (2007). Patent-bibliometric analysis on the Chinese science—technology linkages. *Scientometrics*, *72*(3), 403-425.

Gudivada, V. N., Baeza-Yates, R. A., & Raghavan, V. V. (2015). Big Data: Promises and Problems. *IEEE Computer*, 48(3), 20-23.

Gupta, M., & George, J. F. (2016). Toward the development of a big data analytics capability. *Information & Management*, *53*(8), 1049-1064.

Guzys, D., Dickson-Swift, V., Kenny, A., & Threlkeld, G. (2015). Gadamerian philosophical hermeneutics as a useful methodological framework for the Delphi technique. *International journal of qualitative studies on health and well-being*, *10*(1), 26291.

Hackman, J. R. (1987). The design of work teams. Inj. w. lorsch (ed.), Handbook of organizational behavior (pp. 315-342).

Hackney, R. A., Dooley, P., Levvy, Y., & Parrish, J. (2015). Critical value factors in business intelligence systems implementation success: An empirical analysis of system and information quality. International Conference on Information Systems (ICIS2015-SIGDSA). Texas, USA.

Hair, J. F., Sarstedt, M., Ringle, C. M., & Mena, J. A. (2012). An assessment of the use of partial least squares structural equation modeling in marketing research. *Journal of the academy of marketing science*, *40*(3), 414-433.

Hair, J. F., Ringle, C. M., & Sarstedt, M. (2013). Partial least squares structural equation modeling: Rigorous applications, better results and higher acceptance.

Hair, J. F., Ringle, C. M., & Sarstedt, M. (2011). PLS-SEM: Indeed a silver bullet. *Journal of Marketing theory and Practice*, *19*(2), 139-152.

Hair, J. F., Sarstedt, J., Hopkins, L., & Kuppelwieser, V. (2014). Partial least squares structural equation modeling (PLS-SEM) An emerging tool in business research. *European Business Review*, *26*(2), 106-121.

Hanafin, S. (2004). Review of literature on the Delphi Technique. *Dublin: National Children's Office*.

Haque, A., & Anwar, S. (2012). Linking Top Management Support and IT Infrastructure with Organizational Performance: Mediating Role of Knowledge Application. *Canadian Social Science*, 8(1), 121-129. doi: 10.3968/j.css.1923669720120801.1559

Halaweh, M., & Massry, A. E. (2015). Conceptual model for successful implementation of big data in organizations. *Journal of International Technology and Information Management*, *24*(2), 2.

Hallikainen, P., Kimpimaki, H., & Kivijarvi, H. (2006, January). Supporting the module sequencing decision in the ERP implementation process. In *Hawaii International Conference On System Sciences* (Vol. 39, p. 181). IEEE.

Hanafin, S. (2004). Review of literature on the Delphi Technique. *Dublin: National Children's Office*.

Hartmann, P. M., Zaki, M., Feldmann, N., & Neely, A. (2014). Big data for big business? A taxonomy of data-driven business models used by start-up firms. *A taxonomy of data-driven business models used by start-up firms*.

Hasan, H. M., Lotfollah, F., & Negar, M. (2012). Comprehensive Model of Business Intelligence: a Case Study of Nano's Companies. *Indian Journal of Science and Technology*, *5*(6), 2851-2859.

Havens, T. C., Bezdek, J. C., Leckie, C., Hall, L. O., & Palaniswami, M. (2012). Fuzzy c-means algorithms for very large data. *IEEE Transactions on Fuzzy Systems*, *20*(6), 1130-1146.

Hawking, P., & Sellitto, C. (2010, December). Business Intelligence (BI) critical success factors. In *21st Australian conference on informafion systems* (pp. 1-3).

Haynes, S. N., Nelson, K., & Blaine, D. D. (1999). Psychometric issues in assessment research.

Haynes, S. N., Richard, D., & Kubany, E. S. (1995). Content validity in psychological assessment: a functional approach to concepts and methods. *Psychological assessment*, *7*(3), 238.

Hazen, B. T., Boone, C. A., Ezell, J. D., & Jones-Farmer, L. A. (2014). Data quality for data science, predictive analytics, and big data in supply chain management: An introduction to the problem and suggestions for research and applications. *International Journal of Production Economics*, *154*, 72-80.

He, W., Shen, J., Tian, X., Li, Y., Akula, V., Yan, G., & Tao, R. (2015). Gaining competitive intelligence from social media data: Evidence from two largest retail chains in the world. *Industrial Management & Data Systems*, 115(9), 1622-1636. doi:10.1108/IMDS-03- 2015-0098

Hedges, L. V., & Rhoads, C. (2010). Statistical power analysis. In *International Encyclopedia of Education*. Elsevier Ltd.

Heffner, M., & Sharif, N. (2008). Knowledge fusion for technological innovation in organizations. *Journal of Knowledge Management*, *12*(2), 79-93.

Heiko, A. (2012). Consensus measurement in Delphi studies: review and implications for future quality assurance. *Technological forecasting and social change*, *79*(8), 1525-1536.

Henry, R. M., McCray, G. E., Purvis, R. L., & Roberts, T. L. (2007). Exploiting organizational knowledge in developing IS project cost and schedule estimates: An empirical study. *Information & Management*, *44*(6), 598-612.

Henseler, J., Dijkstra, T. K., Sarstedt, M., Ringle, C. M., Diamantopoulos, A., Straub, D. W., ... & Calantone, R. J. (2014). Common beliefs and reality about PLS: Comments on Rönkkö and Evermann (2013). *Organizational Research Methods*, *17*(2), 182-209.

Henseler, J., Hubona, G., & Ray, P. A. (2016). Using PLS path modeling in new technology research: updated guidelines. *Industrial management & data systems*, *116*(1), 2-20.

Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the academy of marketing science*, *43*(1), 115-135.

High, R. (2000). Important factors in designing statistical power analysis studies. *Computing News, Summer Issue*, 14-15.

Hilbert, M. (2016). Big data for development: A review of promises and challenges. *Development Policy Review, 34*(1), 135-174.

Hochachka, W. M., Fink, D., Hooker, G., Kelling, S., Riedewald, M., Caruana, R., & Ballard, G. (2009). Data-intensive Science: A New Paradigm for Biodiversity Studies. *BioScience*, *59*(7).

Hoe, S. L. (2008). Issues and procedures in adopting structural equation modeling technique. *Journal of applied quantitative methods*, *3*(1), 76-83.

Hoelter, J. W. (1983). The analysis of covariance structures: Goodness-of-fit indices. *Sociological Methods & Research*, *11*(3), 325-344.

Hoser, B., & Nitschke, T. (2010). Questions on ethics for research in the virtually connected world. *Social Networks*, *32*(3), 180-186.

House, R., Javidan, M., Hanges, P., & Dorfman, P. (2002). Understanding cultures and implicit leadership theories across the globe: an introduction to project GLOBE. *Journal of world business*, *37*(1), 3-10.

Howell, D., Windahl, C., & Seidel, R. (2010). A project contingency framework based on uncertainty and its consequences. *International Journal of Project Management*, *28*(3), 256-264.

Hoy, M. B. (2014). Big data: An introduction for librarians. *Medical reference services quarterly*, *33*(3), 320-326.

Howsawi, E. M., Eager, D. M., Bagia, R., & Niebecker, K. D. (2014). The four-level project success framework: application and assessment. *Organisational Project Management*.

Hsu, C. C., & Sandford, B. A. (2007). Minimizing non-response in the Delphi process: How to respond to non-response. *Practical Assessment, Research & Evaluation*, *12*(17), 62-78.

Hu, H., Wen, Y., Chua, T. S., & Li, X. (2014). Toward scalable systems for big data analytics: A technology tutorial. *IEEE access*, *2*, 652-687.

Hu, P. J., Chau, P. Y., Sheng, O. R. L., & Tam, K. Y. (1999). Examining the technology acceptance model using physician acceptance of telemedicine technology. *Journal of management information systems*, *16*(2), 91-112.

Huang, T. C.-K., Liu, C.-C., & Chang, D.-C. (2012). An empirical investigation of factors influencing the adoption of data mining tools. *International Journal of Information Management*, 32(3), 257-270. doi: 10.1016/j.ijinfomgt.2011.11.006

Huemann, M. (2010). Considering Human Resource Management when developing a project-oriented company: Case study of a telecommunication company. *International Journal of Project Management*, *28*(4), 361-369.

Hulland, J. (1999). Use of partial least squares (PLS) in strategic management research: A review of four recent studies. *Strategic management journal*, *20*(2), 195-204.

Hung, S.-Y., Chen, C., & Kuan-Hsiuang, W. (2014). Critical Success Factors for the Implementation of Integrated Healthcare Information Systems Projects: An Organizational Fit Perspective. *Communications of the Association for Information Systems*, 34, 775-796.

Hutka, T. (2009). 21 tips for successful capital projects. Public Management Journal, 91(5), 14.

Hübner-Bloder, G., & Ammenwerth, E. (2009). Key performance indicators to benchmark hospital information systems–a delphi study. *Methods of Information in Medicine*, *48*(06), 508-518.

Hyväri, I. (2006). Success of projects in different organizational conditions. *Project management journal*, *37*(4), 31-41.

Iamratanakul, S., F. Badir, Y., Siengthai, S., & Sukhotu, V. (2014). Indicators of best practices in technology product development projects: Prioritizing critical success factors. *International Journal of Managing Projects in Business*, *7*(4), 602-623.

Ika, L. A. (2009). Project success as a topic in project management journals. *Project Management Journal*, *40*(4), 6-19.

Ika, L. A., Diallo, A., & Thuillier, D. (2012). Critical success factors for World Bank projects: An empirical investigation. *International journal of project management*, *30*(1), 105-116.

Indelicato, G. (2007, February). In case of emergency: Think your project is a disaster? Try these tips from real-life emergency relief teams. *PM Network*, 21(2), 42-44.

Işık, Ö., Jones, M. C., & Sidorova, A. (2013). Business intelligence success: The roles of BI capabilities and decision environments. *Information & Management*, *50*(1), 13-23.

Ives, Z. G., Florescu, D., Friedman, M., Levy, A., & Weld, D. S. (1999, June). An adaptive query execution system for data integration. In *ACM SIGMOD Record* (Vol. 28, No. 2, pp. 299-310). ACM.

Jacobs, A. (2009). The pathologies of big data. *Queue*, *7*(6), 10.

Jafari, A., McGee, P., & Carmean, C. (2006). Managing courses defining learning: What faculty, students, and administrators want. *Educause review*, *41*(4), 50-52.

Jagadish, H. V., Gehrke, J., Labrinidis, A., Papakonstantinou, Y., Patel, J. M., Ramakrishnan, R., & Shahabi, C. (2014). Big data and its technical challenges. *Communications of the ACM*, 57(7), 86-94.

Janssen, M., van der Voort, H., & Wahyudi, A. (2017). Factors influencing big data decision- making quality. *Journal of Business Research*, 70, 338-345.

Jedd, M. (2007). Lifting the Curse-It's a frequent scapegoat, but project management rarely causes project failure. To stop the downward cycle, align projects with culture and strategy. *PM Network*, *21*(3), 62.

Jelinek, M., & Litterer, J. A. (1988). Why OD must become strategic. *Research in organizational change and development*, *2*(1), 135-162.

Jiang, J. J., Klein, G., & Balloun, J. (1996). Ranking of system implementation success factors. *Project Management Journal*, *27*, 49-53.

Ji-fan Ren, S., Fosso Wamba, S., Akter, S., Dubey, R., & Childe, S. J. (2017). Modelling quality dynamics, business value and firm performance in a big data analytics environment. *International Journal of Production Research*, *55*(17), 5011-5026.

Jin, X., Wah, B. W., Cheng, X., & Wang, Y. (2015). Significance and challenges of big data research. *Big Data Research*, *2*(2), 59-64.

Jobs, C. G., Aukers, S. M., & Gilfoil, D. M. (2015). The impact of big data on your firms marketing communications: a framework for understanding the emerging marketing analytics industry. *Academy of Marketing Studies Journal*, *19*(2).

Johnson, J. E. (2012). Big data+ big analytics= big opportunity: big data is dominating the strategy discussion for many financial executives. As these market dynamics continue to evolve, expectations will continue to shift about what should be disclosed, when and to whom. *Financial Executive*, *28*(6), 50-54.

Jones, D. L., Wagstaff, K., Thompson, D. R., D'Addario, L., Navarro, R., Mattmann, C., ... & Rebbapragada, U. (2012, March). Big data challenges for large radio arrays. In *Aerospace Conference, 2012 IEEE* (pp. 1-6). IEEE.

Junqué de Fortuny, E., Martens, D., & Provost, F. (2013). Predictive modeling with big data: is bigger really better?. *Big Data*, *1*(4), 215-226.

Kache, F., & Seuring, S. (2017). Challenges and opportunities of digital information at the intersection of Big Data Analytics and supply chain management. *International Journal of Operations & Production Management*, *37*(1), 10-36.

Kaiser, F. (1970) A second generation little jiffy. *Psychometrika*, 35(4):401–415

Kaiser, H. F., & Rice, J. (1974). Little jiffy, mark IV. *Educational and psychological measurement*, *34*(1), 111-117.

Kaisler, S., Armour, F., Espinosa, J. A., & Money, W. (2013). Big data: Issues and challenges moving forward. In *System sciences (HICSS), 2013 46th Hawaii international conference on* (pp. 995-1004). IEEE.

Kamal, M. M. (2006). IT innovation adoption in the government sector: identifying the critical success factors. *Journal of Enterprise Information Management*, *19*(2), 192-222.

Kamhawi, E. M. (2007). Critical factors for implementation success of ERP systems: an empirical investigation from Bahrain. *International Journal of Enterprise Information Systems (IJEIS)*, *3*(2), 34-49.

Kamioka, T., & Tapanainen, T. (2014). Organizational Use of Big Data and Competitive Advantage-Exploration of Antecedents. In *PACIS* (p. 372).

Katal, A., Wazid, M., & Goudar, R. H. (2013, August). Big data: issues, challenges, tools and good practices. In *Contemporary Computing (IC3), 2013 Sixth International Conference on* (pp. 404-409). IEEE.

Kauffman, R. J., Srivastava, J., & Vayghan, J. (2012). Business and data analytics: New innovations for the management of e-commerce. *Electronic Commerce Research and Applications*, *11*(2), 85-88.

Keller, S. A., Koonin, S. E., & Shipp, S. (2012). Big data and city living–what can it do for us?. *Significance*, *9*(4), 4-7.

Kelly, J., & Kaskade, J. (2013). CIOS & BIG DATA what your IT team wants you to know. *DOI= http://blog. infochimps. com/2013/01/24/cios-big-data*.

Kemp, M. J., & Low, G. C. (2008). ERP innovation implementation model incorporating change management. *Business Process Management Journal*, *14*(2), 228-242.

Kerlinger, F. N., & Lee, H. B. (2000). Survey research. *Foundations of behavioral research*, 599-619.

Kerzner, H. (2003). Strategic planning for a project office. *Project Management Journal*, *34*(2), 13-25.

Ketola, E., & Klockars, M. (1999). Computer-assisted telephone interview (CATI) in primary care. *Family Practice*, *16*(2), 179-183.

Khang, D. B., & Moe, T. L. (2008). Success criteria and factors for international development projects: A life-cycle-based framework. *Project Management Journal*, *39*(1), 72-84.

Khojasteh, N., Ansari, R., & Abadi, H. R. D. (2013). A study of the influencing technological and technical factors successful implementation of business intelligence system in internet service providers companies. *International Journal of Academic Research in Accounting, Finance and Management Sciences*, *3*(2), 125-132.

Kim, M. K., & Park, J. H. (2017). Identifying and prioritizing critical factors for promoting the implementation and usage of big data in healthcare. *Information Development*, *33*(3), 257-269.

Kim, M., Zimmermann, T., DeLine, R., & Begel, A. (2016, May). The emerging role of data scientists on software development teams. In *Proceedings of the 38th International Conference on Software Engineering* (pp. 96-107). ACM.

Kimble, C., & Milolidakis, G. (2015). Big data and business intelligence: Debunking the myths. *Global Business and Organizational Excellence*, *35*(1), 23-34.

Kiron, D., & Shockley, R. (2011). Creating business value with analytics. *MIT Sloan Management Review*, 53(1), 57.

Kisielnicki, J. (2011). The communication system in project teams: Problems of transfer of knowledge and information for the management of IT projects. *Issues in Informing Science and Information Technology*, *8*(unknown), 351-361.

Kitchin, R., Lauriault, T. P., & McArdle, G. (2015). Knowing and governing cities through urban indicators, city benchmarking and real-time dashboards. *Regional Studies, Regional Science*, *2*(1), 6-28.

Kitchin, R., & McArdle, G. (2016). What makes Big Data, Big Data? Exploring the ontological characteristics of 26 datasets. *Big Data & Society*, 3(1).

Knox-Hayes, J. (2013). The spatial and temporal dynamics of value in financialization: Analysis of the infrastructure of carbon markets. *Geoforum*, *50*, 117-128.

Kohnke, O., Wolf, T. R., & Mueller, K. (2011). Managing user acceptance: an empirical investigation in the context of business intelligence standard software. *International Journal of Information Systemsand Change Manag*, *5*(4), 269.

Kolker, E., Stewart, E., & Özdemir, V. (2012). DELSA Global for “big data” and the Bioeconomy: Catalyzing collective innovation. *Industrial Biotechnology*, *8*(4), 176-178.

Koronios, A., Gao, J., & Selle, S. (2014). Big data project success-A meta analysis. In *PACIS* (p. 376).

Koskela, L. J., & Howell, G. (2002). The underlying theory of project management is obsolete. In *Proceedings of the PMI Research Conference* (pp. 293-302). PMI.

Kraemmerand, P., Moller, C., & Boer, H. (2003). ERP implementation: An integrated process of radical change and continuous learning. *Production planning & Control* , 14 (4), 338-348.

Krawatzeck, R., Dinter, B., & Thi, D. A. P. (2015, January). How to make business intelligence agile: The Agile BI actions catalog. In *System Sciences (HICSS), 2015 48th Hawaii International Conference on* (pp. 4762-4771). IEEE.

Krishnan, R., Martin, X., & Noorderhaven, N. G. (2006). When does trust matter to alliance performance?. *Academy of Management journal*, *49*(5), 894-917.

Kubick, W. R. (2012). Big data, information and meaning. *Applied Clinical Trials*, *21*(2), 26.

Kuen, C. W., & Zailani, S. (2012). Critical factors in successful new product development: an empirical study of Malaysian Manufacturing Companies. *International Journal of management*, *29*(2), 429.

Kulkarni, U., & Robles-Flores, J. A. (2013). Development and validation of a BI success model.

Kumar, A., Niu, F., & Ré, C. (2013). Hazy: making it easier to build and maintain big-data analytics. *Communications of the ACM*, *56*(3), 40-49.

Kwon, O., Lee, N., & Shin, B. (2014). Data quality management, data usage experience and acquisition intention of big data analytics. *International Journal of Information Management*, *34*(3), 387-394.

Labrinidis, A., & Jagadish, H. V. (2012). Challenges and opportunities with big data. *Proceedings of the VLDB Endowment*, *5*(12), 2032-2033.

Lane, J. (2012). O privacy, where art thou?: Protecting privacy and confidentiality in an era of big data access. *Chance*, *25*(4), 39-41.

Laney, D. (2001). 3D data management: Controlling data volume, velocity and variety. *META Group Research Note.* 6, 70.

Latonio, M. A. G. (2007). *Exploring the impact of organizational culture on project management: A phenomenological study* (Vol. 68, No. 07).

Laoledchai, Y., Land, L. P. W., & Low, G. (2008). Improving the effectiveness of end-user training outcomes. *ACIS 2008 Proceedings*, 103.

Lautenbach, P., Johnston, K., & Adeniran-Ogundipe, T. (2017). Factors influencing business intelligence and analytics usage extent in South African organisations. *South African Journal of Business Management*, *48*(3), 23-33.

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.

Lazar, J., Feng, J. H., & Hochheiser, H. (2010). Research methods in HCI.

Lazer, D., Kennedy, R., King, G., & Vespignani, A. (2014). The parable of Google Flu: traps in big data analysis. *Science*, *343*(6176), 1203-1205.

Lee, J., Lapira, E., Bagheri, B., & Kao, H. A. (2013). Recent advances and trends in predictive manufacturing systems in big data environment. *Manufacturing Letters*, *1*(1), 38-41.

Lee, J. C., Shiue, Y. C., & Chen, C. Y. (2016). Examining the impacts of organizational culture and top management support of knowledge sharing on the success of software process improvement. *Computers in Human Behavior, 54*, 462-474.

Lee, J. H., Clarke, R. I., & Perti, A. (2015). Empirical evaluation of metadata for video games and interactive media. *Journal of the Association for Information Science and Technology*, *66*(12), 2609-2625.

Lee, P. D., & Hirshfield, M. (2006). Project planning for health care software implementations. *The health care manager*, *25*(4), 310-314.

Li, E. Y. (1997). Perceived importance of information system success factors: A meta analysis of group differences. *Information & management*, *32*(1), 15-28.

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of psychology*.

Liu, G., Wang, E., & Chua, C. (2015). Persuasion and management support for IT projects. *International Journal of Project Management,* 33(6), 1249-1261. doi: 10.1016/j.ijproman.2015.03.009

Loh, T. C., & Koh, S. C. L. (2004). Critical elements for a successful enterprise resource planning implementation in small-and medium-sized enterprises. *International journal of production research*, *42*(17), 3433-3455.

Lohmöller, J. B. (1989). Predictive vs. structural modeling: Pls vs. ml. In *Latent Variable Path Modeling with Partial Least Squares* (pp. 199-226). Physica, Heidelberg.

Lohr, S. (2012). How big data became so big. *New York Times*, *11*.

Loughlin, K. G., & Moore, L. F. (1979). Using Delphi to achieve congruent objectives and activities in a pediatrics department. *Journal of medical education*, *54*(2), 101-106.

Ludwig, B. (1997). Predicting the future: Have you considered using the Delphi methodology. *Journal of extension*, *35*(5), 1-4.

Lukoianova, T., & Rubin, V. L. (2014). *Veracity roadmap: Is big data objective, truthful and credible?.* Advances In Classification Research Online. 10.7152/acro.v24i1.14671.

MacCarthy, B. L., & Atthirawong, W. (2003). Factors affecting location decisions in international operations–a Delphi study. *International Journal of Operations & Production Management*, *23*(7), 794-818.

MacKenzie, S. B., & Podsakoff, P. M. (2012). Common method bias in marketing: Causes, mechanisms, and procedural remedies. *Journal of Retailing*, *88*(4), 542-555.

Maclennan, E., & Van Belle, J.-p. (2014). Factors affecting the organizational adoption of service-oriented architecture (SOA). *Information Systems and eBusiness Management*, 12(1), 71-100. doi: 10.1007/s10257-012-0212-x

MacMillan, I. C., & Day, D. L. (1987). Corporate ventures into industrial markets: Dynamics of aggressive entry. *Journal of business venturing*, *2*(1), 29-39.

Madden, S. (2012). From databases to big data. *IEEE Internet Computing*, (3), 4-6.

Mantelero, A., & Vaciago, G. (2015). Data protection in a big data society. Ideas for a future regulation. *Digital Investigation*, 15, 104-109.

Marbán, O., Menasalvas, E., & Fernández-Baizán, C. (2008). A cost model to estimate the effort of data mining projects (DMCoMo). *Information Systems*, 33(1), 133-150. doi: 10.1016/j.is.2007.07.004

Markus, M. L., & Tanis, C. (2000). The enterprise systems experience-from adoption to success. *Framing the domains of IT research: Glimpsing the future through the past*, *173*, 207-173.

Markus, M. L., Axline, S., Petrie, D., & Tanis, S. C. (2000). Learning from adopters' experiences with ERP: problems encountered and success achieved. *Journal of information technology*, *15*(4), 245-265.

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.

Martin, A. G., & Frick, M. J. (1998). The Delphi Technique: An Informal History of Its Use in Agricultural Education Research since 1984. *Journal of agricultural education*, *39*(1), 73-79.

Martin, T. N., & Huq, Z. (2007). Realigning top management's strategic change actions for ERP implementation: how specializing on just cultural and environmental contextual factors could improve success. *Journal of Change Management*, *7*(2), 121-142.

Martínez-Torres, M. R., & Diaz-Fernandez, M. D. C. (2014). Current issues and research trends on open-source software communities. *Technology Analysis & Strategic Management*, *26*(1), 55-68.

Marvin, H. J., Janssen, E. M., Bouzembrak, Y., Hendriksen, P. J., & Staats, M. (2017). Big data in food safety: An overview. *Critical reviews in food science and nutrition*, *57*(11), 2286-2295.

Mata, J., Portugal, P., & Guimaraes, P. (1995). The survival of new plants: Start-up conditions and post-entry evolution. *International Journal of Industrial Organization*, *13*(4), 459-481.

Mavetera, N., & Kroeze, J. H. (2009). A grounding framework for developing adaptive software products.

Mayfield, C. A., Wingenbach, G. J., & Chalmers, D. R. (2005). Assessing stakeholder needs: Delphi meets the Internet. *Journal of Extension*, *43*(3).

Mäkipää, M. (2003, August). Implementation of Enterprise Resource Planning System–theoretical research approach and empirical evaluation in two cases. In *Proceedings of the 26th information systems research seminar in Scandinavia*.

McAfee, K. (2012). The contradictory logic of global ecosystem services markets. *Development and change*, *43*(1), 105-131.

McAfee, A., Brynjolfsson, E., (2012). Big Data: The Management Revolution. *Harvard Business Review*. 90.

McGrath, J. E., & Brinberg, D. (1983). External validity and the research process: A comment on the Calder/Lynch dialogue. *Journal of Consumer Research*, *10*(1), 115-124.

McHaney, R., White, D., & Heilman, G. E. (2002). Simulation project success and failure: Survey findings. *Simulation & Gaming*, *33*(1), 49-66.

Meijer, E. (2011). The world according to LINQ. *Communications of the ACM*, *54*(10), 45-51.

Melli, G., Wu, X., Beinat, P., Bonchi, F., Cao, L., Duan, R., . . . Zalane, O. (2012). Top 10 data mining case studies. *International Journal of Information Technology & Decision Making*, 11(2), 389-400. doi: 10.1142/S021962201240007X

Menon, S. P., & Hegde, N. P. (2015, January). A survey of tools and applications in big data. In *Intelligent Systems and Control (ISCO), 2015 IEEE 9th International Conference on* (pp. 1-7). IEEE.

Merriam, S. B., & Tisdell, E. J. (2015). *Qualitative research: A guide to design and implementation*. John Wiley & Sons.

Mervis, J. (2012). Agencies rally to tackle big data.

Michael, K., & Miller, K. W. (2013). Big data: New opportunities and new challenges [guest editors' introduction]. *Computer*, *46*(6), 22-24.

Miller, H. E. (2013). Big-data in cloud computing: A taxonomy of risks. *Information Research*, *18*(1), 18-1.

Milosevic, D. Z., & Srivannaboon, S. (2006). A theoretical framework for aligning project management with business strategy. *Project Management Journal*, *37*(3), 98-110.

Mir, F. A., & Pinnington, A. H. (2014). Exploring the value of project management: linking project management performance and project success. *International journal of project management*, *32*(2), 202-217.

Mishra, P., Dangayach, G. S., & Mittal, M. L. (2011). An Ethical approach towards sustainable project Success. *Procedia-Social and Behavioral Sciences*, *25*, 338-344.

Mittal, R. K. (2014). Using data mining techniques for predictive modelling in the retail context. *International Journal of Advanced Research in IT and Engineering,* 3(1), 38-50.

Moohebat, M. R., Jazi, M. D., & Asemi, A. (2011). Evaluation of the ERP implementation at Esfahan steel company based on five critical success factors: a case study. *International Journal of Business and Management*, *6*(5), 236.

Mudzana, T., & Maharaj, M. (2015). Measuring the success of business-intelligence systems in South Africa: An empirical investigation applying the DeLone and McLean Model. *South African Journal of Information Management*, *17*(1), 1-7.

Mudzana, T., & Maharaj, M. (2017). Prioritizing the Factors Influencing the Success of Business Intelligence Systems: A Delphi Study. *Indian Journal of Science and Technology*, *10*(25).

Murad, R. S., & Cavana, R. Y. (2012). Applying the viable system model to ICT project management. *International Journal of Applied Systemic Studies*, *4*(3), 186-205.

Murray, J. P. (2001). Recognizing the responsibility of a failed information technology project as a shared failure. *Information Systems Management*, *18*(2), 25-29.

Muthusamy, S. K., Wheeler, J. V., & Simmons, B. L. (2005). Self-managing work teams: enhancing organizational innovativeness.

Müller, R., & Jugdev, K. (2012). Critical success factors in projects: Pinto, Slevin, and Prescott–the elucidation of project success. *International Journal of Managing Projects in Business*, *5*(4), 757-775.

Müller, R., & Turner, J. R. (2007). Matching the project manager's leadership style to project type. *International journal of project management*, *25*(1), 21-32.

Müller, R., Turner, R., Andersen, E. S., Shao, J., & Kvalnes, Ø. (2014). Ethics, trust, and governance in temporary organizations. *Project Management Journal*, *45*(4), 39-54.

Nagendra, P. B. (2000). Leadership development for the information economy. International Journal of Commerce & Management , 10 (3/4).

Nah, F. F. H., & Delgado, S. (2006). Critical success factors for enterprise resource planning implementation and upgrade. *Journal of Computer Information Systems*, *46*(5), 99-113.

Nah, F. F. H., Islam, Z., & Tan, M. (2007). Empirical assessment of factors influencing success of enterprise resource planning implementations. *Journal of Database Management (JDM)*, *18*(4), 26-50.

Nah, F. F. H., Zuckweiler, K. M., & Lee-Shang Lau, J. (2003). ERP implementation: chief information officers' perceptions of critical success factors. *International journal of Human-computer Interaction*, *16*(1), 5-22.

Nasab, S. S., Selamat, H., & Masrom, M. (2015). A delphi study of the important factors for BI system implementation in the public sector organizations. *Jurnal Teknologi*, *77*(19), 113-120.

Nemec, R. (2011). Business intelligence system success factors in context of DeLone & McLean's information system success assessment model with some application scenarios. In *9th International Conference on Strategic Management and its Support by Information Systems* (pp. 105-111).

Netemeyer, R. G., Bearden, W. O., & Sharma, S. (2003). *Scaling procedures: Issues and applications*. Sage Publications.

Neuman, W. L. (2003). Survey research. *Social Research Methods: Qualitative and Quantitative Approaches. 5th edn. USA: Allyn and Bacon*, 289-290.

Newell, S. (2004). Enhancing cross-project learning. *Engineering Management Journal*, *16*(1), 12-20.

Nguyen, Q., Kuntz, J. R., Näswall, K., & Malinen, S. (2016). Employee resilience and leadership styles: The moderating role of proactive personality and optimism. *New Zealand Journal of Psychology (Online)*, *45*(2), 13.

Nichols, W. (2013). Advertising Analytics 2.0. *Harvard Business Review*, *91*(3), 60-68.

Nimmagadda, S. L., & Dreher, H. V. (2013, July). Big-data integration methodologies for effective management and data mining of petroleum digital ecosystems. In *Digital Ecosystems and Technologies (DEST), 2013 7th IEEE International Conference on* (pp. 148-153). IEEE.

Nixon, P., Harrington, M., & Parker, D. (2012). Leadership performance is significant to project success or failure: a critical analysis. *International Journal of productivity and performance management*, *61*(2), 204-216.

Nonaka, I., & Teece, D. J. (2001). *Managing Industrial Knowledge: New Perspectives on Knowlede-based Firms*. Sage.

Nonaka, I., Toyama, R., & Konno, N. (2000). SECI, Ba and leadership: a unified model of dynamic knowledge creation. *Long range planning*, *33*(1), 5-34.

O'Driscoll, A., Daugelaite, J., & Sleator, R. D. (2013). 'Big data', Hadoop and cloud computing in genomics. *Journal of biomedical informatics*, *46*(5), 774-781.

Oh, Y. K., & Min, J. (2015). The mediating role of popularity rank on the relationship between advertising and in-app purchase sales in mobile application market. *Journal of Applied Business Research*, *31*(4), 1311.

Ohata, M., & Kumar, A. (2012). Big data: a boon to business intelligence. *Financial Executive*, *28*(7), 63-65.

Ohlhorst, F. J. (2012). *Big data analytics: turning big data into big money*. John Wiley & Sons.

Ofori, D. F. (2013). Project management practices and critical success factors–a developing country perspective. *International Journal of Business and Management*, *8*(21), 14.

Ojiako, U., Johansen, E., & Greenwood, D. (2008). A qualitative re-construction of project measurement criteria. *Industrial Management & Data Systems*, *108*(3), 405-417.

Olbrich, S., Poeppelbuss, J., & Niehaves, B. (2011). BI systems managers' perception of critical contextual success factors: A Delphi study.

Olsson, N. O. (2006). Management of flexibility in projects. *International Journal of Project Management*, *24*(1), 66-74.

Olszak, C. M., & Ziemba, E. (2012). Critical success factors for implementing business intelligence systems in small and medium enterprises on the example of upper Silesia, Poland. *Interdisciplinary Journal of Information, Knowledge, and Management*, *7*(12), 129-150.

Olugbode, M., Elbeltagi, I., Simmons, M., & Biss, T. (2008). The Effect of Information Systems on Firm Performance and Profitability Using a Case-Study Approach. *Electronic Journal of Information Systems Evaluation*, *11*(1).

Opresnik, D., & Taisch, M. (2015). The value of big data in servitization. *International Journal of Production Economics*, *165*, 174-184.

Palanisamy, R., Verville, J., Bernadas, C., & Taskin, N. (2010). An empirical study on the influences on the acquisition of enterprise software decisions. *Journal of Enterprise Information Management,* 23(5), 610-639. doi: 10.1108/17410391011083065

Palcic, I., & Buchmeister, B. (2012). Project success in Slovenian companies. *DAAAM International Scientific Book*, 53-65.

Palinkas, L. A., Horwitz, S. M., Green, C. A., Wisdom, J. P., Duan, N., & Hoagwood, K. (2015). Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. *Administration and Policy in Mental Health and Mental Health Services Research*, *42*(5), 533-544.

Pandit, N. R. (1996). The creation of theory: A recent application of the grounded theory method. *The qualitative report*, *2*(4), 1-15.

Passannante, M. R., Gallagher, C. T., & Reichman, L. B. (1994). Preventive therapy for contacts of multidrug-resistant tuberculosis: a Delphi survey. *Chest*, *106*(2), 431-434.

Pasukeviciute, I., & Roe, M. (2001). The politics of oil in Lithuania: strategies after transition. *Energy policy*, *29*(5), 383-397.

Patton, M. Q. (2015). Qualitative research & evaluation methods: Integrating theory and practice.

Peslak, A., & Stanton, M. (2007). Information technology team achievement: An analysis of success factors and development of a team success model (TSM). *Team Performance Management: An International Journal*, *13*(1/2), 21-33.

Peslak, A. R., Subramanian, G. H., & Clayton, G. E. (2008). The phases of ERP software implementation and maintenance: A model for predicting preferred ERP use. *Journal of Computer Information Systems*, *48*(2), 25-33.

Peterson, N. A., Speer, P. W., & McMillan, D. W. (2008). Validation of a brief sense of community scale: Confirmation of the principal theory of sense of community. *Journal of community psychology*, *36*(1), 61-73.

Pham, Q. T., Mai, T. K., Misra, S., Crawford, B., & Soto, R. (2016, July). Critical success factors for implementing business intelligence system: Empirical study in vietnam. In *International Conference on Computational Science and Its Applications* (pp. 567-584). Springer, Cham.

Phillips, J. J. (2009). *Show Me the Money: How to Determine Roi in People, Projects, and Programs: Easyread Large Bold Edition*. ReadHowYouWant. com.

Pinto, J. K. (2014). Project management, governance, and the normalization of deviance. *International Journal of Project Management*, *32*(3), 376-387.

Pinto, J. K., & Prescott, J. E. (1988). Variations in critical success factors over the stages in the project life cycle. *Journal of management*, *14*(1), 5-18.

Pinto, J. K., & Slevin, D. P. (1987). Critical factors in successful project implementation. *IEEE transactions on engineering management*, (1), 22-27.

Pinto, J. K., & Slevin, D. P. (1989). Critical success factors in R&D projects. *Research-technology management*, *32*(1), 31-35.

Plano Clark, V. L., Schumacher, K., West, C., Edrington, J., Dunn, L. B., Harzstark, A., ... & Miaskowski, C. (2013). Practices for embedding an interpretive qualitative approach within a randomized clinical trial. *Journal of Mixed Methods Research*, *7*(3), 219-242.

Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of applied psychology*, *88*(5), 879.

Podsakoff, P. M., & Organ, D. W. (1986). Self-reports in organizational research: Problems and prospects. *Journal of management*, *12*(4), 531-544.

Podsakoff, P. M., & Todor, W. D. (1985). Relationships between leader reward and punishment behavior and group processes and productivity. *Journal of Management*, *11*(1), 55-73.

Pope-Ruark, R. (2015). Introducing agile project management strategies in technical and professional communication courses. *Journal of Business and Technical Communication*, *29*(1), 112-133.

Popovič, A., Hackney, R., Tassabehji, R., & Castelli, M. (2018). The impact of big data analytics on firms' high value business performance. *Information Systems Frontiers*, *20*(2), 209-222.

Popovič, A., Turk, T., & Jaklič, J. (2010). Conceptual model of business value of business intelligence systems. *Management: Journal of Contemporary Management Issues*, *15*(1), 5-30.

Pospiech, M., & Felden, C. (2016, January). Big Data--A Theory Model. In *2016 49th Hawaii International Conference on System Sciences (HICSS)* (pp. 5012-5021). IEEE.

Powell, W. W., & Snellman, K. (2004). The knowledge economy. *Annu. Rev. Sociol.*, *30*, 199-220.

Provost, F., & Fawcett, T. (2013). Data science and its relationship to big data and data-driven decision making. *Big data*, *1*(1), 51-59.

Puklavec, B., Oliveira, T., & Popovič, A. (2014). Unpacking Business Intelligence Systems Adoption Determinants: An Exploratory Study of Small and Medium Enterprises. *Economic & Business Review*, *16*(2).

Punathambekar, A., & Kavada, A. (2015). Debating Big Data. *Media, Culture & Society*.

Purdam, K. (2016). Task-based learning approaches for supporting the development of social science researchers' critical data skills. *International Journal of Social Research Methodology*, 19(2), 257-267. doi:10.1080/13645579.2015.1102453

Putnam, J. W., Spiegel, A. N., & Bruininks, R. H. (1995). Future directions in education and inclusion of students with disabilities: A Delphi investigation. *Exceptional Children*, *61*(6), 553-576.

Rai, A., Lang, S. S., & Welker, R. B. (2002). Assessing the validity of IS success models: An empirical test and theoretical analysis. *Information systems research*, *13*(1), 50-69.

Rajan, K. (2015). Materials informatics: The materials "gene" and big data. *Annual Review of Materials Research*, 45, 153-169. doi:10.1146/annurev-matsci-070214-021132

Rajpurohit, A. (2013, October). Big data for business managers—Bridging the gap between potential and value. In *Big Data, 2013 IEEE International Conference on* (pp. 29-31). IEEE.

Ramayah, T., Lee, J. W. C., & In, J. B. C. (2011). Network collaboration and performance in the tourism sector. *Service Business*, *5*(4), 411.

Ravasan, A. Z., & Savoji, S. R. (2014). An Investigation of BI Implementation Critical Success Factors in Iranian Context. *International Journal of Business Intelligence Research (IJBIR)*, *5*(3), 41-57.

Reel, J. S. (1999). Critical success factors in software projects. *IEEE software*, (3), 18-23.

Rebentrost, P., Mohseni, M., & Lloyd, S. (2014). Quantum support vector machine for big data classification. *Physical review letters*, *113*(13), 130503.

Reddi, V. J., Lee, B. C., Chilimbi, T., & Vaid, K. (2011). Mobile processors for energy-efficient web search. *ACM Transactions on Computer Systems (TOCS)*, *29*(3), 9.

Rees, N., Rapport, F., & Snooks, H. (2015). Perceptions of paramedics and emergency staff about the care they provide to people who self-harm: constructivist metasynthesis of the qualitative literature. *Journal of psychosomatic research*, *78*(6), 529-535.

Reimsbach-Kounatze, C. (2015), "The Proliferation of "Big Data" and Implications for Official Statistics and Statistical Agencies: A Preliminary Analysis", *OECD Digital Economy Papers*, No. 245, OECD Publishing, Paris,https://doi.org/10.1787/5js7t9wqzvg8-en.

Reinartz, W., Haenlein, M., & Henseler, J. (2009). An empirical comparison of the efficacy of covariance-based and variance-based SEM. *International Journal of research in Marketing*, *26*(4), 332-344.

Richins, M. L., & Dawson, S. (1992). A consumer values orientation for materialism and its measurement: Scale development and validation. *Journal of consumer research*, *19*(3), 303-316.

Rigdon, E. E. (2014). Rethinking partial least squares path modeling: breaking chains and forging ahead. *Long Range Planning*, *47*(3), 161-167.

Riggins, F. J., & Wamba, S. F. (2015). Research directions on the adoption, usage, and impact of the internet of things through the use of big data analytics. In *System Sciences (HICSS), 2015 48th Hawaii International Conference on* (pp. 1531-1540). IEEE.

Robson, C. (2002). Real world research: A resource for social scientists and practitioner-researchers.

Rockart, J. F. (1979). Chief executives define their own data needs. *Harvard business review*, *57*(2), 81-93.

Rodríguez-Mañas, L., Féart, C., Mann, G., Viña, J., Chatterji, S., Chodzko-Zajko, W., ... & Scuteri, A. (2012). Searching for an operational definition of frailty: a Delphi method based consensus statement. The frailty operative definition-consensus conference project. *Journals of Gerontology Series A: Biomedical Sciences and Medical Sciences*, *68*(1), 62-67.

Rolstadås, A., Tommelein, I., Morten Schiefloe, P., & Ballard, G. (2014). Understanding project success through analysis of project management approach. *International journal of managing projects in business*, *7*(4), 638-660.

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

Rubinstein, I. S. (2014). Voter privacy in the age of big data. *Wis. L. Rev.*, 861.

Russom, P. (2013). Managing big data. *TDWI Best Practices Report, TDWI Research*, 1-40.

Sadovskyi, O., Engel, T., Heininger, R., Böhm, M., & Krcmar, H. (2014). Analysis of big data enabled business models using a value chain perspective. *Multikonferenz Wirtschaftsinformatik (MKWI 2014)*, 1126-1137.

Saldaña, J. (2009). An introduction to codes and coding. *The coding manual for qualitative researchers*, *3*.

Salmasi, M. K., Talebpour, A., & Homayounvala, E. (2016). Identification and classification of organizational level competencies for BI success. *Journal of Intelligence Studies in Business*, *6*(2).

Saltz, J. S. (2015). The need for new processes, methodologies and tools to support big data teams and improve big data project effectiveness. In *Big Data (Big Data), 2015 IEEE International Conference on* (pp. 2066-2071). IEEE.

Saltz, J. S., & Shamshurin, I. (2016). Big data team process methodologies: A literature review and the identification of key factors for a project's success. In *Big Data (Big Data), 2016 IEEE International Conference on* (pp. 2872-2879). IEEE.

Saltz, J., Shamshurin, I., & Connors, C. (2016). A Framework for Describing Big Data Projects. In *International Conference on Business Information Systems* (pp. 183-195). Springer, Cham.

Saran, C. (2012). Almost a third of BI projects fail to deliver on business objectives. *Computer Weekly*.

Sarkis, J., & Sundarraj, R. P. (2003). Managing large-scale global enterprise resource planning systems: a case study at Texas Instruments. *International Journal of Information Management*, *23*(5), 431-442.

Sauer, G., Holman, J., Lazar, J., Hochheiser, H., & Feng, J. (2010). Accessible privacy and security: a universally usable human-interaction proof tool. *Universal Access in the Information Society*, *9*(3), 239-248.

Sauser, B. J., Reilly, R. R., & Shenhar, A. J. (2009). Why projects fail? How contingency theory can provide new insights–A comparative analysis of NASA's Mars Climate Orbiter loss. *International Journal of Project Management*, *27*(7), 665-679.

Schachter, E. P. (2004). Identity configurations: A new perspective on identity formation in contemporary society. *Journal of personality*, *72*(1), 167-200.

Schadt, E. E., Linderman, M. D., Sorenson, J., Lee, L., & Nolan, G. P. (2010). Computational solutions to large-scale data management and analysis. *Nature reviews genetics*, *11*(9), 647.

Scott, S. L., Blocker, A. W., Bonassi, F. V., Chipman, H. A., George, E. I., & McCulloch, R. E. (2016). Bayes and big data: The consensus Monte Carlo algorithm. *International Journal of Management Science and Engineering Management*, *11*(2), 78-88.

Scott, S. G., & Bruce, R. A. (1994). Determinants of innovative behavior: A path model of individual innovation in the workplace. *Academy of management journal*, *37*(3), 580-607.

Seay, C., Agrawal, R., Kadadi, A., & Barel, Y. (2015, April). Using hadoop on the mainframe: A big solution for the challenges of big data. In *Information Technology-New Generations (ITNG), 2015 12th International Conference on* (pp. 765-769). IEEE.

Seddon, P. B. (1997). A respecification and extension of the DeLone and McLean model of IS success. *Information systems research*, *8*(3), 240-253.

Seddon, P. B., Calvert, C., & Yang, S. (2010). A multi-project model of key factors affecting organizational benefits from enterprise systems. *MIS quarterly*, *34*(2), 305-328.

Seddon, J. J., & Currie, W. L. (2017). A model for unpacking big data analytics in high-frequency trading. *Journal of Business Research*, *70*, 300-307.

Segars, A. H., & Grover, V. (1998). Strategic information systems planning success: an investigation of the construct and its measurement. *MIS quarterly*, 139-163.

Sekaran, U., & Bougie, R. (2010). Theoretical framework In theoretical framework and hypothesis development. *Research methods for business: A skill building approach*, *80*.

Shao, M. G. (2006). Development of project manager selection tool based on project manager competency. *Master's Abstracts International*, 45 (02)

Shatat, A. S. (2015). Critical success factors in enterprise resource planning (ERP) system implementation: An exploratory study in Oman. *Electronic Journal of Information Systems Evaluation*, *18*(1), 36.

Shenhar, A. J., & Wideman, R. M. (2000). Optimizing project success by matching PM style with project type. In *Project Management Forum*.

Shenhar, A. J., Dvir, D., Levy, O., & Maltz, A. C. (2001). Project success: A multidimensional strategic concept. *Long Range Planning*, 34(6), 699-725.

Shimp, T. A., & Sharma, S. (1987). Consumer ethnocentrism: Construction and validation of the CETSCALE. *Journal of marketing research*, 280-289.

Shin, D. H., & Choi, M. J. (2015). Ecological views of big data: Perspectives and issues. *Telematics and Informatics*, *32*(2), 311-320.

Shinn, G. C., Wingenbach, G. J., Briers, G. E., Lindner, J. R., & Baker, M. (2009). Forecasting doctoral–level content in international agricultural and extension education–2010: Viewpoint of fifteen engaged scholars. *Journal of International Agricultural and Extension Education*, *16*(1), 57-71.

Sidawi, B. (2012). The impact of social interaction and communications on innovation in the architectural design studio. *Buildings*, *2*(3), 203-217.

Siemens, G., & Long, P. (2011). Penetrating the fog: Analytics in learning and education. *EDUCAUSE review*, *46*(5), 30.

Sim, J. (2014). Consolidation of success factors in data mining projects. *GSTF Journal on Computing (JoC)*, 4(1), 66-73.

Simon, H. A. (2013). *Administrative behavior*. Simon and Schuster.

Simonin, B. L. (1999). Ambiguity and the process of knowledge transfer in strategic alliances. *Strategic management journal*, *20*(7), 595-623.

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.

Skinner, R., Nelson, R. R., Chin, W. W., & Land, L. (2015). The Delphi Method Research Strategy in Studies of Information Systems. *CAIS*, *37*, 2.

Smith, M., Szongott, C., Henne, B., & Von Voigt, G. (2012, June). Big data privacy issues in public social media. In *Digital Ecosystems Technologies (DEST), 2012 6th IEEE International Conference on* (pp. 1-6). IEEE.

Snow, C.P. (1966). Government Science and Public Policy. *Science*. Vol. 151, (650-653).

Snyder-Halpern, R. (2001). Indicators of organizational readiness for clinical information technology/systems innovation: a Delphi study. *International journal of medical informatics*, *63*(3), 179-204.

Soares, S. (2012). *Big data governance: an emerging imperative*. Mc Press.

Sobek, M., Cleveland, L., Flood, S., Kelly Hall, P., King, M. L., Ruggles, S., & Schroeder, M. (2011). Big data: large-scale historical infrastructure from the Minnesota Population Center. *Historical methods, 44*(2), 61-68.

Soja, P. (2008). Difficulties in enterprise system implementation in emerging economies: Insights from an exploratory study in Poland. *Information Technology for Development*, *14*(1), 31-51.

Solow, B. L. (1987). Capitalism and slavery in the exceedingly long run. *The Journal of Interdisciplinary History*, *17*(4), 711-737.

Somers, T. M., & Nelson, K. G. (2004). A taxonomy of players and activities across the ERP project life cycle. *Information & Management*, *41*(3), 257-278.

Sookhak, M., Gani, A., Khan, M. K., & Buyya, R. (2017). Dynamic remote data auditing for securing big data storage in cloud computing. Information Sciences, 380, 101-116. doi:10.1016/j.ins.2015.09.004

Söderholm, A. (2008). Project management of unexpected events. *International Journal of Project Management*, *26*(1), 80-86.

Sparks, B. H., & McCann, J. T. (2015). Factors influencing business intelligence system use in decision making and organisational performance. *International Journal of Sustainable Strategic Management*, *5*(1), 31-54.

Spender, J. C. (1996). Making knowledge the basis of a dynamic theory of the firm. *Strategic management journal*, *17*(S2), 45-62.

Standish Group (1994), The CHAOS Report, The Standish Group International, Inc, Boston, MA.

Stankovic, D., Nikolic, V., Djordjevic, M., & Cao, D. B. (2013). A survey study of critical success factors in agile software projects in former Yugoslavia IT companies. *Journal of Systems and Software*, *86*(6), 1663-1678.

Starns, J., & Odom, C. (2006). Using knowledge management principles to solve organizational performance problems. *Vine*, *36*(2), 186-198.

Sterba, K. R., DeVellis, R. F., Lewis, M. A., Baucom, D. H., Jordan, J. M., & DeVellis, B. M.(2007). Developing and testing a measure of dyadic efficacy for married women with rheumatoid arthritis and their spouses. *Arthritis & Rheumatism (Arthritis Care & Research)*, *57*(2), 294-302.

Sternad, S., & Bobek, S. (2006). Factors which have fatal influence on ERP implementation on Slovenian organizations. *Journal of information and organizational sciences*, *30*(2), 279-293.

Stitt-Gohdes, W. L., & Crews, T. B. (2004). The Delphi technique: A research strategy for career and technical education. *Journal of Career and Technical Education*, *20*(2), 55-67.

Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. *Journal of the royal statistical society. Series B (Methodological)*, 111-147.

Strauss, A., & Corbin, J. (1994). Grounded theory methodology. *Handbook of qualitative research*, *17*, 273-85.

Strawn, G. O. (2012). Scientific Research: How Many Paradigms?. *Educause Review*, *47*(3), 26.

Sudhakar, G. P. (2012). A model of critical success factors for software projects. *Journal of Enterprise Information Management*, *25*(6), 537-558.

Sumsion, T. (1998). The Delphi technique: an adaptive research tool. *British Journal of Occupational Therapy*, *61*(4), 153-156.

Sun, F., Huang, G. B., Wu, Q. J., Song, S., & Wunsch II, D. C. (2017). Efficient and rapid machine learning algorithms for big data and dynamic varying systems. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, *47*(10), 2625-2626.

Sutphin, H. D., & Camp, W. G. (1990). A model for building consensus on the applications of microcomputers in agricultural education. *Journal of Vocational Education Research*, *15*(3), 65-79.

Tabladillo, M. (2008). Why use Microsoft Data Mining. *Microsoft SQL Server*.

Tankard, C. (2012). Big data security. *Network security*, *2012*(7), 5-8.

Tarhini, A., Ammar, H., Tarhini, T., & Masa'deh, R. E. (2015). Analysis of the Critical Success Factors for Enterprise Resource Planning Implementation from Stakeholders' Perspective: A Systematic Review. *International Business Research*, 8(4), 25-40.

Taylor, S., & Todd, P. A. (1995a). Understanding information technology usage: A test of competing models. *Information systems research*, *6*(2), 144-176.

Taylor, S., & Todd, P. (1995b). Decomposition and crossover effects in the theory of planned behavior: A study of consumer adoption intentions. *International journal of research in marketing*, *12*(2), 137-155.

Teece, D. J. (2014). The foundations of enterprise performance: Dynamic and ordinary capabilities in an (economic) theory of firms. *Academy of management perspectives*, *28*(4), 328-352.

Teece, D. J. (2015). Intangible assets and a theory of heterogeneous firms. In *Intangibles, market failure and innovation performance* (pp. 217-239). Springer, Cham.

Teece, D. J., Pisano, G., & Shuen, A. (1997). Dynamic capabilities and strategic management. *Strategic management journal*, *18*(7), 509-533.

Tenenhaus, M., Vinzi, V. E., Chatelin, Y. M., & Lauro, C. (2005). PLS path modeling. *Computational statistics & data analysis*, *48*(1), 159-205.

Thomas, G., & Fernández, W. (2008). Success in IT projects: A matter of definition?. *International journal of project management*, *26*(7), 733-742.

Thompson, R. L., Higgins, C. A., & Howell, J. M. (1991). Personal computing: toward a conceptual model of utilization. *MIS quarterly*, 125-143.

Tona, O., Carlsson, S., & Eom, S. (2012). An empirical test of DeLone and McLean's information system success model in a public organization.

Tonidandel, S., King, E. B., & Cortina, J. M. (Eds.). (2015). *Big data at work: The data science revolution and organizational psychology*. Routledge.

Tormay, P. (2015). Big data in pharmaceutical R&D: Creating a sustainable R&D engine. Pharmaceutical medicine, 29(2), 87-92.

Turner, J. R., & Müller, R. (2004). Communication and co-operation on projects between the project owner as principal and the project manager as agent. *European Management Journal*, *22*(3), 327-336.

Turner, J. R., & Müller, R. (2005). The project manager's leadership style as a success factor on projects: A literature review. *Project management journal*, *36*(2), 49-61.

Uğur, N. G., & Turan, A. H. Strategies for BI acceptance: challenges and solutions. *PressAcademia Procedia*, *7*(1), 237-240.

Urbach, N., & Ahlemann, F. (2010). Structural equation modeling in information systems research using partial least squares. *Journal of Information technology theory and application*, *11*(2), 5-40.

VanVoorhis, C. W., & Morgan, B. L. (2007). Understanding power and rules of thumb for determining sample sizes. *Tutorials in Quantitative Methods for Psychology*, *3*(2), 43-50.

Venkatesh, V., & Davis, F. D. (2000). A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management science*, *46*(2), 186-204.

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS quarterly*, 425-478.

Vera-Baquero, A., Colomo-Palacios, R., & Molloy, O. (2013). Business process analytics using a big data approach. *IT Professional*, *15*(6), 29-35.

Verma, V. (1995). The human aspects of project management: organizing projects for success. Project Management Institute.

Villars, R. L., Olofson, C. W., & Eastwood, M. (2011). Big data: What it is and why you should care. *White Paper, IDC, 14*, 1-14.

Visinescu, L. L., Jones, M. C., & Sidorova, A. (2017). Improving decision quality: the role of business intelligence. *Journal of Computer Information Systems, 57*(1), 58-66.

Wagner, J. (2012). International trade and firm performance: a survey of empirical studies since 2006. *Review of World Economics*, *148*(2), 235-267.

Wamba, S., 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. *Int. J. Prod. Econ.* http://dx.doi.org/10.1016/j.ijpe.2014.12.031

Wang, C. H. (2016). A novel approach to conduct the importance-satisfaction analysis for acquiring typical user groups in business-intelligence systems. *Computers in Human Behavior, 54*, 673-681.

Wang, Y., & Byrd, T. A. (2017). Business analytics-enabled decision-making effectiveness through knowledge absorptive capacity in health care. *Journal of Knowledge Management*, *21*(3), 517-539.

Wang, Y., & Hajli, N. (2017). Exploring the path to big data analytics success in healthcare. *Journal of Business Research*, *70*, 287-299.

Wang, E. T., Klein, G., & Jiang, J. J. (2006). ERP misfit: country of origin and organizational factors. *Journal of Management Information Systems*, *23*(1), 263-292.

Wang, J., & Wang, X. (2012). *Structural equation modeling: Applications using Mplus*. John Wiley & Sons.

Wearne, S. (2008). Stakeholders in excellence in teaching and learning of project management. *International Journal of Project Management, 26*(3), 326-328.

Wenrich, K., & Ahmad, N. (2009). Lessons learned during a decade of ERP experience: A case study. *International Journal of Enterprise Information Systems (IJEIS)*, *5*(1), 55-73.

White, L. (2009). Challenge of research ethics committees to the nature of operations research. *Ethics and Operations Research*, 37(6), 1083-1088. doi:10.1016/j.omega.2008.12.003.

Wieder, B., Ossimitz, M., & Chamoni, P. (2012). The impact of business intelligence tools on performance: a user satisfaction paradox?.

Wielki, J. (2013, September). Implementation of the big data concept in organizations-possibilities, impediments and challenges. In *Computer Science and Information Systems (FedCSIS), 2013 Federated Conference on* (pp. 985-989). IEEE.

Willcocks, L. P. (1999). Evaluating the outcomes of information systems plans.

Willcocks, L., & Lester, S. (1996). Beyond the IT productivity paradox. *European Management Journal*, *14*(3), 279-290.

Wixom, B., Ariyachandra, T., Douglas, D. E., Goul, M., Gupta, B., Iyer, L. S., ... & Turetken, O. (2014). The current state of business intelligence in academia: The arrival of big data. *CAIS*, *34*, 1.

Wixom, B. H., & Watson, H. J. (2001). An empirical investigation of the factors affecting data warehousing success. *MIS quarterly*, 17-41.

Wong, B., & Tein, D. (2003). Critical success factors for ERP projects. In *Project Management conference*. AIPM.

Wood, J. (2008). Effective project management. *Health Facilities Management*, 21(6), 49.

Worrell, J. L., Di Gangi, P. M., & Bush, A. A. (2013). Exploring the use of the Delphi method in accounting information systems research. *International Journal of Accounting Information Systems*, *14*(3), 193-208.

Wu, X., Zhu, X., Wu, G. Q., & Ding, W. (2014). Data mining with big data. *IEEE transactions on knowledge and data engineering*, *26*(1), 97-107.

Wuest, J. (1995). Feminist grounded theory: An exploration of the congruency and tensions between two traditions in knowledge discovery. *Qualitative health research*, *5*(1), 125-137.

Xu, G., & Gutiérrez, J. A. (2006). An exploratory study of killer applications and critical success factors in m-commerce. *Journal of Electronic Commerce in Organizations (JECO)*, *4*(3), 63-79.

Xu, W., Huang, R., Zhang, H., El-Khamra, Y., & Walling, D. (2016). Empowering R with high performance computing resources for big data analytics. In *Conquering Big Data with High Performance Computing* (pp. 191-217). Springer, Cham.

Xu, Z., Liu, Y., Mei, L., Hu, C., & Chen, L. (2015). Semantic based representing and organizing surveillance big data using video structural description technology. *Journal of Systems and Software*, *102*, 217-225.

Yang, S., & Seddon, P. (2004). Benefits and key project success factors from enterprise systems implementations: lessons from Sapphire 2003. *ACIS 2004 Proceedings*, 27.

Yeoh, W., & Koronios, A. (2010). Critical success factors for business intelligence systems. *Journal of computer information systems*, *50*(3), 23-32.

Yeoh, W., & Popovič, A. (2016). Extending the understanding of critical success factors for implementing business intelligence systems. *Journal of the Association for Information Science and Technology*, *67*(1), 134-147.

Yew Wong, K. (2005). Critical success factors for implementing knowledge management in small and medium enterprises. *Industrial management & Data systems*, *105*(3), 261-279.

Yildirim, C., & Correia, A. P. (2015). Exploring the dimensions of nomophobia: Development and validation of a self-reported questionnaire. *Computers in Human Behavior*, *49*, 130-137.

Yin, S., & Kaynak, O. (2015). Big data for modern industry: challenges and trends [point of view]. *Proceedings of the IEEE, 103*(2), 143-146.

Yogev, N., Even, A., & Fink, L. (2013). How business intelligence creates value: An empirical investigation. *International Journal of Business Intelligence Research (IJBIR)*, *4*(3), 16-31.

Yoon, K., Hoogduin, L., & Zhang, L. (2015). Big Data as complementary audit evidence. *Accounting Horizons*, *29*(2), 431-438.

Young, R., & Jordan, E. (2008). Top management support: Mantra or necessity? *International Journal of Project Management*, 26(7), 713-725. doi: 10.1016/j.ijproman.2008.06.001

Young, R., & Poon, S. (2013). Top management support—almost always necessary and sometimes sufficient for success: Findings from a fuzzy set analysis. *International Journal of Project Management*, 31(7), 943-957. doi: 10.1016/j.ijproman.2012.11.013

Young, S. D. (2014). Behavioral insights on big data: using social media for predicting biomedical outcomes. *Trends in microbiology*, *22*(11), 601-602.

Zaichkowsky, J. L. (1985). Measuring the involvement construct. *Journal of consumer research*, *12*(3), 341-352.

Zhang, Z., Lee, M. K., Huang, P., Zhang, L., & Huang, X. (2005). A framework of ERP systems implementation success in China: An empirical study. *International Journal of Production Economics*, *98*(1), 56-80.

Zhao, R., Liu, Y., Zhang, N., & Huang, T. (2017). An optimization model for green supply chain management by using a big data analytic approach. *Journal of Cleaner Production*, *142*, 1085-1097.

Zheng, L., Zeng, C., Li, L., Jiang, Y., Xue, W., Li, J., . . . Wang, P. (2014). *Applying data mining techniques to address critical process optimization needs in advanced manufacturing*. Paper presented at the Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining, New York, New York, USA.

Zhou, L., Pan, S., Wang, J., & Vasilakos, A. V. (2017). Machine learning on big data: Opportunities and challenges. *Neurocomputing*, *237*, 350-361.

Zikopoulos, P., & Eaton, C. (2011). *Understanding big data: Analytics for enterprise class hadoop and streaming data*. McGraw-Hill Osborne Media.

Zott, C., & Amit, R. (2007). Business model design and the performance of entrepreneurial firms. *Organization science*, *18*(2), 181-199.

Thesis

Adamala, S., & Cidrin, L. (2011). *Key success factors in business intelligence*. Student thesis. Blekinge Institute of Technology, School of Management

Agirre Perez, I. (2007). Stochastic project scheduling system: Implications for risk management. *Dissertation Abstracts International, 68*(2).

Akbıyık, A. (2012). *Uzaktan Eğitim Ortamlarinda Sosyal Yazilim Kullaniminin Kabulünü Etkileyen Faktörlerin Belirlenmesine Yönelik Bir Çalişma.* (Doctoral Dissertation). Sakarya University.

Baladi, I. W. (2007). An empirical analysis of perceived value of virtual versus traditional project management practice. *Dissertation Abstracts International*, 68 (9).

Cook, B. W. (2004). Measuring the value of success in project management organizations. *Dissertation Abstracts International*, 67 (5).

Cornell University. (2016). *A Guide to Conducting Systematic Reviews: Steps in a Systematic Review*. http://guides.library.cornell.edu/c.php?g=459012&p=3142201

Delisle, C. L. G. (2001). Success and communication in virtual project teams. Dissertation *Abstracts International*, 62 (12), 4242.

Fan, H. (2007). Leveraging operation data for intelligent decision support in construction equipment management. *Dissertation Abstracts International*, 68 (10).

Field, A. (2009). *Discovering statistics using SPSS*. Sage publications.

Ildefonso, G. A. (2007). "Case Studies of Mergers and Acquisitions: Best Practices for Technology Transfer." *Dissertation Abstracts International*, 60 (1).

Jones, B. (2007). Factors affecting the full and successful implementation of new technology supporting national security: Analysis of the implementation of the Single Mobility System. *Dissertation Abstracts International*, 68 (09). (UMI

Liemi, P. C. (2004). A study of the relationship between improving the management of projects and the use of knowledge management. *Dissertation Abstracts International*, 43 (03), 1608B.

Nasr, E. B. (2004). An integrated project planning and control system approach for measuring project performance. *Dissertation Abstracts International*, 66 (03).

Nieder, N. (2016). *Effective Big Data Mangement: a development of critical sucess factors and an analysis of firms' capabilities in the automotive industry* (Doctoral dissertation).

Pinto, J. K. (1986). *Project implementation: a determination of its critical success factors, moderators and their relative importance across the project life cycle* (Doctoral dissertation). University of Pittsburgh.

Ramsey, J. W. (2009). *Identifying Entry-level Skills Expected by Agricultural Industry Experts and Determining Teachers' Perceptions on Whether They Are Being Learned Through Students' Participation in the Supervised Argricultural Experience Component of the Secondary Agricultural Education Program: a Two-panel Delphi Study* (Doctoral dissertation). Oklahoma State University.

Sauser, B. J. (2005). Assessing NASA strategic project leadership in the era of 'better, faster, cheaper'. *Dissertation Abstracts International*, 66 (5).

Sidenko, S. (2006). *Information technology project management: project management maturity and its effect on project success.* (unpublished doctoral dissertation). Concordia University.

Westlund, S. G. (2007). Retaining talent: Assessing relationships among project leadership styles, software developer job satisfaction, and turnover intentions. *Dissertation Abstracts International*, 68 (11).

Wu, W. W. (2006). IT personnel sourcing decisions in IT projects. *Dissertation Abstracts International*, 67 (1).

Zhao, F. (2007). An empirical study of enterprise system upgrades. *Dissertation Abstracts International*, 68 (4).

Internet Sources

Allouche, G. (2014). *How Big Data can Save Health Care*. http://www.innovationexcellence.com/blog/2014/11/14/how-big-data-can-save-health-care/

Ang, L. E. (2009). *Critical factors for achieving data mining success*. (unpublished MBA Thesis), Universiti Sains Malaysia, Malaysia. http://eprints.usm.my/25482/1/CRITICAL_FACTORS_FOR_ACHIEVING_DA TA.pdf

BI Intelligence Estimates. (2015). *Why IoT, big data & smart farming are the future of agriculture*. http://uk.businessinsider.com/internet-of-things-smart-agriculture-2016-10

Burg, N. (2014). How Big Data Will Help Save Healthcare. *Forbes Magazine*, *10*. https://www.forbes.com/sites/castlight/2014/11/10/how-big-data-will-help-save-healthcare/#6efe8a6b1f9e

Cebr (2012). *The Value of Big Data and the Internet of Things to the UK Economy*. https://www.sas.com/content/dam/SAS/en_gb/doc/analystreport/cebr-value-of-big-data.pdf

DBTalks. (2016). *Big Data: The Future Has Arrived*. http://www.dbtalks.com/article/big-data-future-has-come/

Forrester, 2012. *The Big Deal About Big Data For Customer Engagement Business: Leaders Must Lead Big Data Initiatives To Derive Value*. https://www.forrester.com/report/The+Big+Deal+About+Big+Data+For+Customer+Engagement/-/E-RES72241

Gartner. (2017). *Gartner Says Business Intelligence and Analytics Leaders Must Focus on Mindsets and Culture to Kick Start Advanced Analytics*. https://www.gartner.com/newsroom/id/3130017

Grand View Research. (2016). *Big Data Market Size, Share Forecast, Industry Research Report, 2025*. https://www.grandviewresearch.com/industry-analysis/big-data-industry?utm_source=Pressrelease&utm_medium=referral&utm_campaign=abnewswire_05oct&utm_content=content

Hass, K. (2006). *The five deadly sins of project management.* https://www.powermag.com/the-five-deadly-sins-of-project-management/

IDC. (2013). *IDC Forecast: Big data technology and services to hit $32.4 billion in 2017*. https://hostingjournalist.com/cloud-hosting/idc-forecast-big-data-technology-and-services-to-hit-32-4-billion-in-2017/

IDC. (2015). *Double-Digit Growth Forecast for the Worldwide Big Data and Business Analytics Market Through 2020 Led by Banking and Manufacturing*

Investments, *According to IDC* https://www.idc.com/url.do?url=/includes/pdf_download.jsp?containerId=prUS41826116&position=51

MGI. (2012). *Big Data: The next frontier for innovation, competition, and productivity*. https://www.mckinsey.com/~/media/McKinsey/Business%20Functions/McKinsey%20Digital/Our%20Insights/Big%20data%20The%20next%20frontier%20for%20innovation/MGI_big_data_exec_summary.ashx

PMI. (2015). *Executive engagement: The role of the sponsor*. https://www.pmi.org/business-solutions/white-papers/executive-engagement-sponsor-role

Standish Group (2010). *Chaos.* https://standishgroup.com/newsroom/modernization.php

TechAmerica Foundation. (2012). *Demystifying Big Data: A practical guide to transforming the business of government.* http://www.techamerica.org/Docs/fileManager.cfm?f=techamerica-bigdatareport-final.pdf

Techrepublich. (2017). *85% of big data projects fail, but your developers can help yours succeed.* https://www.techrepublic.com/article/85-of-big-data-projects-fail-but-your-developers-can-help-yours-succeed/

U.S. General Accounting Office. (1997). *Department of Energy: Improving management of major system acquisitions.* http://www.gao.gov/archive/1997/rc97092t.pdf

Vass, L. (2016, 28 June). *Terabyte terror: It takes special databases to lasso the Internet of Things*. Ars Technica. https://arstechnica.com/information-technology/2016/06/building-databases-for-the-internet-of-data-spewing-things/

Woodward, K. (2007*). Best practices in project management*. http://www.faulkner.com/showcase/faccts.htm

APPENDICES

Appendix 1: Ethics Committee Approval (1/2)

Evrak Tarih ve Sayısı: 08/03/2018-E.11668

BENN4NE09

T.C.
SAKARYA ÜNİVERSİTESİ REKTÖRLÜĞÜ
Etik Kurulu

Sayı :61923333/050.03/
Konu :81/02 Prof.Dr.Aykut Hamit TURAN

Sayın Doç.Dr. Aykut Hamit TURAN

İlgi : Aykut Hamit TURAN 28/02/2018 tarihli ve 0 sayılı yazı

Üniversitemiz Etik Kurulu Başkanlığının **01.03.2018** tarihli ve **80** sayılı toplantısında alınan **"2"** nolu karar örneği ekte sunulmuştur.
Bilgilerinizi arz ederim.

Prof.Dr. Haluk SELVİ
Etik Kurulu Başkanı

2- Prof.Dr.Aykut Hamit TURAN'ın "Büyük Verinin Organizasyonel Etkileri: Kritik Başarı Faktörleri Üzerine Bir Araştırma" başlıklı çalışması görüşmeye açıldı.
Yapılan görüşmeler sonunda; Prof.Dr.Aykut Hamit TURAN'ın "Büyük Verinin Organizasyonel Etkileri: Kritik Başarı Faktörleri Üzerine Bir Araştırma" başlıklı çalışmasının Etik açıdan uygun olduğuna oy birliği ile karar verildi.

Güvenli Elektronik İmzalı Aslı ile Aynıdır.
24/05/2018

Mesut Ramazan EKİCİ
Fakülte Sekreteri

Etik Kurulu Esentepe Kampüsü 54187 Serdivan SAKARYA / KEP Adresi: sakaryauniversitesi@hs01.kep.tr
Tel:0264 295 50 00 Faks:0264 295 50 31
E-Posta :ozelkalem@sakarya.edu.tr Elektronik Ağ :www.sakarya.edu.tr

Bu belge 5070 sayılı Elektronik İmza Kanununun 5. Maddesi gereğince güvenli elektronik imza ile imzalanmıştır.

Appendix 1: Ethics Committee Approval (2/2)

T.C.
SAKARYA ÜNİVERSİTESİ
ETİK KURULU TOPLANTISI TOPLANTI KATILIM LİSTESİ

Toplantı Adı	: Etik Kurulu Toplantısı		
Toplantı Tarihi	: 01.03.2018	Toplantı No	: 81
Toplantı Yeri	: REKTÖRLÜK BİNASI TOPLANTI ODASI		
Toplantı Başlama Saati	: 14:00	Toplantı Bitiş Saati	: -

ADI-SOYADI	GÖREVİ	İMZA
Prof. Dr. Haluk SELVİ	Fen-Edebiyat Fakültesi	
Prof. Dr. Davut DURSUN	Siyasal Bilgiler Fakültesi	
Prof. Dr. MUSTAFA ALTINDİŞ	Tıp Fakültesi	Katılmadı
Prof. Dr. Abdulvahit İMAMOĞLU	İlahiyat Fakültesi	
Prof. Dr. Ahmet ALP	Mühendislik Fakültesi	
Prof. Dr. Sima NART	İşletme Fakültesi	
Prof. Dr. Cuma BİNDAL	Mühendislik fakültesi	
Prof. Dr. Ayşe ÜSTÜN	Sanat Tasarım ve Mimarlık Fakültesi	
Prof. Dr. Mustafa YILMAZLAR	Eğitim Fakültesi	

Appendix 2: Critical Success Factors Scale

BÜYÜK VERİ PROJELERİNDE KRİTİK BAŞARI FAKTÖRLERİ

Değerli Katılımcı,

Bu anket büyük veri projelerinin başarıya ulaşmasında kritik düzeyde öneme sahip başarı faktörlerinin belirlenmesi amacıyla hazırlanmıştır. Elde edilen veriler, katılımcıların kimlikleri bilinmeden toplu olarak değerlendirilecek ve sadece bilimsel amaçlarla kullanılacaktır.

BÜYÜK VERİ: Standart sistemlerle depolanması, yönetilmesi, analiz edilmesi mümkün olmayan ve yüksek hacim, çeşitlilik ve yenilenme hızı özelliklerine sahip olan veri kümeleridir.

Kaç yıldır büyük veri projelerinde çalışmaktasınız? ☐0-3 yıl ☐4-6 yıl ☐7-9 yıl ☐10 yıl ve üzeri
Katılımcı daha önce hiç büyük veri projesinde yer almadıysa lütfen anketi <u>sonlandırınız.</u>

Anket ifadelerinde;

PROJE; büyük veri projesini ifade etmektedir.

KRİTİK; uygulanmadığı takdirde projenin başarısızlığına sebep olabilecek, yüksek öneme sahip anlamındadır.

Soruları cevaplarken lütfen yer aldığınız EN SON büyük veri projesini değerlendirin.

	Lütfen aşağıdaki ifadelere katılım düzeyinizi belirtiniz. (1: Kesinlikle katılmıyorum – 7: Kesinlikle katılıyorum)	**1**	**2**	**3**	**4**	**5**	**6**	**7**
1	İşletmenin proje ihtiyaç duyulacak kaynaklara (teknoloji, insan vb.) yatırım yapma imkânı olması kritik öneme sahiptir.							
2	Üst yönetimin projeyi desteklemesi kritik öneme sahiptir.							
3	Projenin çözüm bulmayı hedeflediği iş probleminin doğru tanımlanması kritik öneme sahiptir.							
4	Proje amacının ve kapsamının doğru tanımlanması kritik öneme sahiptir.							
5	Proje kapsamının işletmenin ihtiyaçlarıyla örtüşmesi kritik öneme sahiptir.							
6	Projenin iş amaçlarıyla paralellik ve uygunluk göstermesi kritik öneme sahiptir.							
7	İşletmenin amaç ve hedeflerine ulaşmada büyük verinin stratejik konumunun belirlenmesi kritik öneme sahiptir.							
8	Büyük verinin işletmenin iş süreçlerinde ve karar alma süreçlerinde önemli rol oynaması kritik öneme sahiptir.							
9	İşletmenin esnek IT altyapısına sahip olması kritik öneme sahiptir.							
10	Proje ekibi liderinin yönetsel yeteneklere sahip olması kritik öneme sahiptir.							

11	Proje ekibinde ilgili her departmandan çalışanlar yer alması kritik öneme sahiptir.							
12	Proje ekibindeki kişilerin analitik düşünme yeteneğine sahip olması kritik öneme sahiptir.							
13	Proje ekibindeki kişilerin gerekli teknik yeteneklere sahip olması kritik öneme sahiptir.							
14	İşletmenin proje ihtiyaçlarına uygun personel istihdamı kritik öneme sahiptir.							
15	Proje ekibinin büyük veri konusunda eğitimli olması kritik öneme sahiptir.							
16	Proje kapsamında iş tanımlarının doğru yapılması kritik öneme sahiptir.							
17	Projeye ayrılan kaynakların (insan, teknoloji, para vb.) doğru dağıtılması kritik öneme sahiptir.							
18	Proje kapsamında hangi uygulama ve geliştirme araçlarının (hardware, software, Hadoop, Python vb.) kullanılacağına yönelik stratejilerin belirlenmiş olması kritik öneme sahiptir.							
19	Proje takviminin açık ve net olması kritik öneme sahiptir.							
20	Projede teknoloji-iş-insan dengesinin sağlanması kritik öneme sahiptir.							
21	İşletmenin proje kapsamında teknoloji insan görev ve örgüt yapısındaki değişimleri yönetebilmesi kritik öneme sahiptir.							
22	Projenin kısa zamanda fayda sağlayarak ilerlemesi (kilometre taşlarının hızlı erişilebilir olması) kritik öneme sahiptir.							
23	Projenin kısa sürede sonuçlanabilecek ve uygulamaya konabilecek olması kritik öneme sahiptir.							
24	Ekipteki kişilerin birbirleriyle sağlıklı iletişim kurabilmeleri kritik öneme sahiptir.							
25	Ekiptekilerin proje süresince dokümantasyon yapmak konusunda hassasiyet göstermesi kritik öneme sahiptir.							
26	Proje ekibinin sistemi kullanacak olan kişilerle sağlıklı etkileşimi kritik öneme sahiptir.							
27	Proje kapsamında kullanılan verinin kaliteli (eksiksiz, tutarlı, doğru, uygun vb.) olması kritik öneme sahiptir.							
28	Proje kapsamında güncel analiz araçlarının kullanılması kritik öneme sahiptir.							
29	Projede eski ve yeni veri tabanlarının entegrasyonunun sağlanabilmesi kritik öneme sahiptir.							
30	Proje süresince gerek duyulan işletme içi/dışı veri kaynaklarına erişilebilmesi kritik öneme sahiptir.							

31	Veri yönetimi ve denetim faaliyetlerinin sorunsuz sağlanması kritik öneme sahiptir.							
32	Projenin ölçülebilir çıktılarının olması kritik öneme sahiptir.							
33	İşletmenin bilişim altyapısının gelecekteki ihtiyaçlar da dikkate alınarak kurulması kritik öneme sahiptir.							
34	Proje hedeflerine başarıyla ulaşılmıştır.							
35	Projenin tüm paydaşları tatmin olmuştur.							
36	Proje bütçe hedefini tutturmuştur.							
37	Proje kalite hedefine ulaşmıştır.							
38	Proje takvimine uygun şekilde tamamlanmıştır.							
39	Projenin başarılı olduğuna inanıyorum.							

Unvanınız: ☐Bilgi Sistemleri Grubu Başkanı (CIO) ☐Genel Müdür ☐Genel Müdür Yrd.

☐Yönetici ☐Yönetici Yrd. ☐Müdür. ☐Müdür Yrd.

☐Uzman ☐Uzman Yrd. ☐Şef ☐Şef Yrd.

☐Koordinatör ☐Koordinatör Yrd. ☐Sorumlu ☐Eleman

İşletmenizin faaliyet gösterdiği sektör:

☐Eğitim ☐Finans ☐Otomotiv ☐Enerji ☐Bilişim Teknolojileri ☐Gıda

☐Ağaç işleri ☐İnşaat ☐Kimya-Plastik ☐Sağlık ☐Elektrik-elektronik ☐Perakende

☐Tekstil ☐Medya İletişim ☐Diğer: belirtiniz__________

İşletmede çalışan toplam personel sayısı: ☐0-9 ☐10-49 ☐50-249 ☐250-500 ☐500 üzeri

Bilgi İşlem biriminde çalışan toplam kişi sayısı: _______

Bilgi İşlem biriminde LİSANS mezunu olan kişi sayısı: _______

Bilgi İşlem biriminde LİSANS ÜSTÜ mezunu olan kişi sayısı: ______

Eğitim düzeyiniz: ☐Lise ☐Ön Lisans ☐Lisans ☐Yüksek Lisans ☐Doktora

Bilişim alanında çalışma süreniz (yıl): _______

Büyük veri üzerine çalışma süreniz (yıl): ☐0-3 yıl ☐4-6 yıl ☐7-9 yıl ☐10 yıl ve üzeri

Cinsiyetiniz: ☐Kadın ☐Erkek

Yaşınız: ☐18-25 ☐26-35 ☐36-45 ☐46-55 ☐56 ve üzeri

CURRICULUM VITAE

Naciye Güliz Uğur was born in İstanbul in 1986. She graduated from VKV Koç Özel Lisesi in 2004 and the Business Informatics Department of Marmara University in 2009. After graduation, she worked for 5 years in private sector. In 2013, she was accepted as research assistant by the Department of Management Information Systems at Sakarya University. She received master's degree in Management Information Systems in 2015 with a thesis entitled "Determining the factors affecting college students' acceptance of mobile applications: A case study of Sakarya University" and was enrolled in the Doctor of Philosophy Program of the department in the same year under supervision of Prof. Dr. Aykut Hamit Turan. She still works as a research assistant at the same department.