2021

Effective use of data analytics and its impact on business performance within small-to-medium-sized businesses

Alfonso Berumen

Pepperdine University

Graziadio School of Business

EFFECTIVE USE OF DATA ANALYTICS AND ITS IMPACT ON BUSINESS PERFORMANCE WITHIN SMALL-TO-MEDIUM-SIZED BUSINESSES

A dissertation submitted in partial fulfilment

of the requirements for the degree of

DOCTOR OF BUSINESS ADMINISTRATION

by

Alfonso Berumen

August, 2021

John Mooney, Ph.D. – Dissertation Chair

This dissertation, written by

Alfonso Berumen

under the guidance of a Dissertation Committee and approved by its members, has been submitted to and accepted by the Pepperdine Graziadio Business School in partial fulfilment of the requirements for the degree of

DOCTOR OF BUSINESS ADMINISTRATION

Doctoral Dissertation Committee:

John Mooney, Ph.D., Supervisor and Chairperson

Mark Chun, Ph.D., Secondary Advisor

Paul Tallon, Ph.D., External Reviewer

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

DEDICATION

This dissertation is dedicated to my family including my mother, Margaret Berumen, father, Alfonso Berumen, and sister, Michelle Berumen. Without their support, the completion of the dissertation would not have been possible.

ACKNOWLEDGMENTS

I would like to acknowledge my committee, particularly Dr. John Mooney, my dissertation chair, Dr. Mark Chun, my secondary advisor, and Dr. Paul Tallon, my external reviewer. Their expertise and guidance have been instrumental to the research presented. Additionally, I would like to thank Dr. Craig Everett for his tireless efforts in recruitment of research participants. Further, I would like to thank the Pepperdine Graziadio Business School DBA faculty. Finally, as a member of the Graziadio School adjunct faculty, I would like to thank the staff and my students that I have worked with during the completion of this dissertation.

VITA

EDUCATION

Doctor of Business Administration, Pepperdine University, Graziadio Business School, Malibu, CA, August 2021

Master of Science, Predictive Analytics, Northwestern University, Chicago, IL, December 2017

Master of Business Administration, University of California at Irvine, Paul Merage School of Business, Irvine, CA, December 2014

Bachelor of Arts, Economics, Occidental College, Los Angeles, CA, May 2008

PROFESSIONAL EMPLOYMENT

Los Angeles Data Analytics LLC

October 2018-Present: Owner, Data Scientist, & Independent Consultant

Resolution Economics LLC

May 2008 to September 2018: Senior Manager; Manager; Senior Consultant; Consultant

ACADEMIC EMPLOYMENT

Pepperdine University, Graziadio Business School

October 2019-Present: Instructor of Decision Sciences & Information Systems and Technology Management; Adjunct Faculty; MS in Business Analytics Graduate Program Committee Member

University of California, Los Angeles – Extension

April 2018-June 2020: Instructor, Data Analytics & Management

PROFESSIONAL MEMBERSHIP

The Institute for Operations Research and the Management Sciences (informs)

Association for Information Systems (AIS)

ABSTRACT

Business use of data analytics and its potential impact on firm performance have become topics of deep interest within both the business practitioner and academic communities. While previous research has demonstrated relationships between data analytics and firm performance in larger firms, there is limited research on whether and how data analytics is used within and impacts Small-to-Medium-sized Business (SMB) settings. Given the preponderance of SMBs within the US economy, and their contribution to employment and economic activity, it is important for SMB owners to understand what management practices lead to effective use of data analytics that in turn impacts SMB performance. Drawing upon the Resource-Based View (RBV) of the firm and prior empirical research on practices within large firms, this dissertation identifies the resources that are needed to form a Data Analytics Capability (DAC) and examines the relationship between the maturity of DACs and the extent of business value realized. The research model was tested using Partial Least Squares-Structural Equation Modelling (PLS-SEM) analysis of survey data gathered from a sample of 300 SMB firms in the US, complemented with qualitative interviews of SMB owners. The results provide evidence that a more developed DAC can lead to higher Data Analytics Business Value across business functions.

Keywords: business analytics, small-to-medium-sized business, business analytics capabilities, analytics business value, PLS

CHAPTER 1: INTRODUCTION

There has been significant attention given to the use of data analytics (DA) to positively impact firm performance by both the information systems (IS) academic community and business practitioners. Studies of the use of data analytics to generate business value are not a recent phenomenon. Davenport (2009) stated, “Analytics and decision automation are among the most powerful tools for improving decision making. A growing number of firms are embracing the former both strategically and tactically, building competitive strategies around their analytical capabilities and making decisions on the basis of data and analytics” (p. 5). Kiron et al. (2012) described a survey from more than 2,500 business executives in which 67% of respondents reported that “using analytics has created at least a moderate competitive advantage for them” (p. 51). More broadly, the application of data analytics for business can be described as business analytics (BA). There has been a more recent focus on how the use of big data analytics (BDA) can impact firm performance. For example, Grover et al. (2018) described BDA as the “application of statistical, processing, and analytics techniques to big data for advancing business” (p. 390).

While IS scholars and professionals have focused on BDA to generate a competitive advantage and drive business value at large corporations, it should also be of interest to the business community whether and how DA and BDA can be leveraged within small, private firms. There are different perspectives and qualifications for what categorizes a small, medium, or large business in the United States. However, a small-to-medium-sized business (SMB) is typically defined as a firm with less than 500 employees. Though individually small in employment, on aggregate these firms are significant contributors to the overall economy. According to the U.S. Small Business Administration, SMBs account for approximately 44% of

U.S. Economic Activity, create two-thirds of net new jobs, and are the driving force behind U.S. innovation and competitiveness. Yet, there is limited understanding of whether DA can be impactful for SMBs, and, if so, what practices are associated with realizing higher levels of business value from analytics capabilities. More specifically, it has been well established that simple accumulation of data and/or adoption of Information Technology (IT) alone does not specifically lead to a firm realizing value (Gupta & George, 2016). Previous research on large corporations that have adopted analytics have identified some common characteristics and factors associated with the development of analytics capabilities that can lead to a perceived impact on firm performance. SMBs and their owners and managers, who often possess less resources than larger firms, should have comparable insights available that will allow them to properly navigate their data analytics strategy.

Adoption and familiarity with data and software in this setting requires careful consideration as smaller firms and their ownership may exhibit more variation in the level of use. In fact, it was revealed through the pilot phase of this dissertation research that a separate mechanism might be required to understand the sophistication of data utilized to capture different frequency levels of reliance on big data versus more traditional data types. Indeed, given that SMBs have not been a priority for big data researchers, it is important to not just focus on BDA applications but also understand the level of frequency with which data that might be classified as ‘big’ is used in an SMB setting. Consistent with the stream of big data research (BDR), the objective of this research is to unearth what factors facilitate the relationship between effective DA use and business value since investment in technologies alone is unlikely to lead to optimal business benefits. Thus, this dissertation research study seeks to fill the gap in understanding the potential business benefits that can be achieved through the effective adoption

and use of DA by SMBs, and more specifically what practices and capabilities are associated with achieving higher levels of business value from DA within SMB settings.

Research Problem Addressed

The purpose of this research is to study the use and exploitation of DA by SMBs in the U.S., including the impact of big data. Further, I seek to determine the technical and organizational capabilities that contribute to the impact of DA on business performance, defined as the business value (BV) of DA, notated as DABV throughout. Finally, I seek to identify the factors that contribute to the development of DA capabilities that in turn contribute to the realization of DABV. To provide insight into the nature and development of analytics capabilities within SMBs, it is appropriate to consider the importance of SMBs in the U.S.

The Importance of SMBs to the U.S. Economy

Within the US economy, SMBs constitute a significant portion of employment and economic activity. Traditionally, a measure used to qualify a firm as small-to-medium sized is the number of people the firm employs. In the U.S., the threshold for SMBs is typically fewer than 500 employees. For example, a recent report produced by Facebook in partnership with the Small Business Roundtable showed that in a survey of approximately 86,000 SMB owners they applied a cut-off of fewer than 500 employees (Facebook, 2021). Additionally, this cut-off has been used in U.S. Census Bureau statistics reporting.

According to Statistics of U.S. Businesses (SUSB, 2017), of the over 5.97 million firms in the U.S., 99.7% had fewer than 500 employees. Total annual sales, receipts, or value of shipments was approximately $373.7 billion; firms with fewer than 500 employees contributed over 35% or approximately $133.19 billion of sales, receipts, or value of shipments (U.S. Census Bureau, 2017). Additionally, SMBs are important to the U.S. economy as these firms employed

over 60.56 million people in 2017 which corresponded to over 47% of employment (U.S. Census Bureau, 2017). Firms with fewer than 500 employees also contributed approximately 40% of annual payroll in the U.S. (U.S. Census Bureau, 2017).

To provide insights into the nature and development of analytics capabilities within SMBs, it is appropriate to first consider some definitions related to DA and its use by businesses in general.

Introduction to SMBs and DA Use

Llave (2019) conducted an extensive literature review of business intelligence and analytics (BI&A) within SMBs which yielded the finding of limited researcher attention to BI&A in a smaller firm setting. Llave (2019) identified 78 articles and found few had more than 30 citations. Additionally, the research method used was dominated by case studies and surveys, with 29 without a clearly stated methodology or alternatively deficient by empirical standards (Llave, 2019). Several sources included by Llave (2019) were those in which the authors focused on adoption and maturity levels of SMBs. However, none appear to be focused on the U.S., consider capabilities, or the contribution to business value. Notably, Llave (2019) argued that SMBs are different than larger enterprises due to their limited IT and financial resources as well as their ownership and operating procedures. Importantly, "*[s]everal studies focused on the components of BI&A* [emphasis added] (Gupta & George, 2016; Mikalef et al., 2017), but they did not address the SME context. Therefore, further research should address the development of additional reference models for BI&A components in SMEs" (Llave, 2019, p. 31). The studies provide the foundation for DA capabilities leveraged in this dissertation.

Lee (2017) reported that with DA, data goes from low value in its original form to a high value strategic asset and, as such, professionals should, "assess the benefits and costs of

collecting and/or generating big data, choose high-value data sources, and build analytics capable of providing value-added information to managers" (p. 295). Lee (2017) discussed the positive impacts of big data on large firms citing examples from organizations who have generated value through personalization marketing, better pricing, cost reduction, and improved customer service. Lee (2017) also identified challenges such as data quality, data security, privacy, data management, and shortage of qualified talent, and noted the investment justification as a challenge citing big data projects could be characterized as having a high risk of failure and investment costs of such projects are considered irreversible. In the context of SMBs, this is a significant business consideration as investment in resources would surely be required to develop a data analytics capability (DAC).

Other researchers have studied the use of IS within SMB contexts. For example, Cragg et al. (2011) relied on resource-based theory to study IS competences within SMBs. Unsurprisingly, the motivation for research was the lack of SMB-focused literature: "The literature indicated that firms possessed many types of IS resources and that many competences were needed for IS success. However, as much of this was for large organizations, the following research question drove our research: What IS competences are relevant to SMBs [SMEs]?" (Cragg et al., 2011, p. 7). The authors argued, "When compared with large enterprises, SMBs [SMEs] usually have a simpler structure with less specialised tasks and with poor human, financial, and material resources" (Cragg et al., 2011, p. 3). Cragg et al. (2011) further argued:

> SMBs [SMEs] are often resource poor, and resource based theory indicates that they will need different competences to cope with scarce resources. They may also have to rely more on external resources and thus a different set of competences are required, particularly externally focused ones. Furthermore,

> organisation theory indicates that SMBs [SMEs] have a flatter/simpler structure and thus, internal coordination is less of an issue as there is close proximity between all staff, including owners and employees. (p. 3)

Cragg et al. (2011) sought to generally understand what IS competences are important for smaller firms to develop but also noted the possibility that some SMBs will possess more IS competences than others and some may require a greater range of competences based on individualized business goals. However, it may be difficult to generalize any of these findings in the context of DA given that many authors have suggested that there are specific resources required to build a DAC.

At the very least, SMBs are understudied despite having some unique organizational characteristics and circumstances when compared to larger firms. Also, given the need to provide clarity when conducting research, I intend to pay special attention to the context of how DA and BDA are defined with respect to SMBs. Thus, the first element of the research problem addressed is the focus on effective DA use within SMBs.

SMB Analytics Maturity, Data Analytics Resources, and Capabilities

SMBs may often have limited financial resources. Thus, DA projects put firm owners in a difficult position as the financial risk associated with successful implementation for a SMB may be much higher than that of a larger firm. While there may be several shared factors (i.e., the need to invest in infrastructure, leadership, and talent) that may similarly challenge both large firms and SMBs, the fact that SMBs are characteristically different in a number of facets has been widely agreed upon by business and IS researchers.

Indeed, previous authors have acknowledged the need to differentiate the factors that explain adoption, use, and benefits realization: "In fact, research suggests that the value of a

technology is affected by the actual degree of usage of individuals (Baura et al., 2010); if the innovation is not in use by any people or processes, then the organization realizes little or no value from its adoption" (Hazen et al., 2012, p. 126). For example, newly adopted innovations, such as DA, can be considered as special projects compartmentalized to a certain individual or organizational unit and fail to be routinized by the organization if not acknowledged as being important to the organization as whole (Hazen et al., 2012). This is consistent with the previous discussion of DA capabilities which require alignment of several components and practices including a data-driven culture.

Shah et al. (2017) highlight the role of big data within small-to-medium sized firms and call for future research to describe the current state of firms and identify traits firms must possess to become more data-driven. Thus, it is important to reveal insights on the variation in DA use by SMBs that may or may not lead to comparable business value gains that have been demonstrated as achievable by larger firms. Finally, while there has been some mention of traits and characteristics and generally capabilities of firms that have been able to leverage BA to create value, an obvious additional consideration is to identify the core capabilities that explain value or whose absence might explain the lack of value. Again, this is consistent with the research goal of understanding what resources or practices lead to effective use of DA to generate value within SMBs so that there is a clearer pathway for decision-makers. In other words, what should an SMB owner invest in given information and patterns of historical success from SMBs who have been able to see an impact on business value?

With respect to SMBs, yet another consideration is the level of DA adoption, DA adoption challenges, and DA maturity within SMBs. Indeed, in studying adoption including theoretical models, factors that influence adoption, and research challenges, Baig et al. (2019)

argued that organizational factors include cost of adoption or "the financial resources available in adopting new technology" and firm size as "Number of resources is directly related to the size of the firm. As organizational size increases, more resources, tools, technical expertise will be required to maintain the big data" (pp. 9-10). In a similar vein, Agrawal (2017) studied the organizational challenges of BDA adoption based on a survey of 106 firms in China and India. Agrawal (2017) found "that the cost of software, hardware, consultancy support, installation and integration are obstacles to BDA adoption. Since bigger firms have greater resources and knowledge to assimilate BDA and also the economies of scale to derive maximum benefit (Gibbs et al., 2004), so firm size has a positive effect on BDA adoption" (p. 7).

Coleman et al. (2016) attempted to identify barriers to SMB adoption of BDA and developed a big data maturity model with a focus on European firms. In this setting, Coleman et al. (2016) defined BA as the following: "the totality of data-based inference methodology used for the objective of analyzing, predicting and controlling processes in business and industry" (pp. 2151-2152) and argued it includes descriptive analytics including graphics, predictive analytics such as forecasting, and prescriptive analytics or the process of translating descriptive and predictive analytics into business action. Yet, Coleman et al. (2016) argued that BDA adoption for SMBs is expected to be behind that of larger corporations. They highlighted that SMBs could face a number of challenges including lack of understanding of BDA, lack of business cases or case studies, lack of suitable talent or in-house data analytic expertise, shortage of effective and affordable consulting and BA services, lack of financial resources, and business model to name a few. More importantly, they argued the need to measure the maturity of an organization first: "The effective implementation of big data programmes in organisations *requires a preliminary assessment of their maturity regarding the strategic use of data in the daily processes* [emphasis

added]" (Coleman et al., 2016, p. 2158). They argued that a maturity model can be evaluated with respect to the following: business strategy, data management, existence of specialized people and analytical skills, technological infrastructure, level of enterprise adoption, leadership and corporate culture, and data governance.

Unsurprisingly, the elements described by Coleman et al. (2016) also appear as themes and factors presented in more empirically established research models which focus on data analytics capabilities. Consistent with the previous discussion on lack of empirical research, Coleman et al. (2016) argued that further research needs to be SMB-focused so that there is a clearer path for SMBs to becoming more analytical. While I do not intend to study adoption and use of DA, previous research suggests that the size of the firm can impact the maturity and sophistication of data analytics use, thereby justifying the need to study smaller firms separately and understand how SMB management perceive these concepts.

Accordingly, one should contemplate the size of the business and what considerations, such as cost or alternatively the allocation of available resources, might stymie a SMB from either pursing data analytics projects or failing to effectively realize value from data analytics projects. Thus, the second element of significance for this dissertation is to understand the presence of DA capabilities that may define different levels of maturity for realizing value.

Insights from BDA on Value Creation

While SMBs have been studied by business researchers in several disciplines, including how these firms adopt and leverage emerging IT resources, there has been limited research on how DA can be used effectively within SMBs to enhance business performance. In addition, BDA is a relatively new sub-category within the broader fields of BA, Business Intelligence (BI), and Business Intelligence Systems (BIS). For example, as part of the development of a

research framework for big data and strategy, Mikalef et al (2016) aggregated big data and BA in the context of its impact on business decision-making:

> Nevertheless, since *big data and business analytics* [emphasis added] are a relatively new technological and business paradigm, there is little research support on how to effectively manage them and use them in the most effective manner. Early studies have shown the benefits of using big data in different contexts, yet, there is a lack of theoretically driven research on how to utilize these solutions in order to gain a competitive advantage. (p. 6)

Chen and Nath (2017) argued BI and BA have been used interchangeably in the literature and adopted BI as representing "technologies and techniques that focus on using a consistent set of metrics to measure past performance to guide business planning" while BA was described as going beyond BI and considered as "an umbrella term to describe all three main types of analytics: descriptive, predictive, and prescriptive" (p. 62). Additionally, Puklavec et al. (2018) studied and framed BIS as "quality information in well-designed data stores, coupled with software tools that provide users timely access, effective analysis and intuitive presentation of the right information, enabling them to take the right actions or make the right decision" and argued BIS can "enable enhancements in firms' strategic planning, business processes, improvements of performance, and building of competitive advantage" (p. 1). Similarly with respect to SMBs, they argued that BIS research has been primarily focused on large-sized firms. Thus, there are some useful insights and research trends pertaining to DA that may be embedded within the literature on BDA, BA, BI, and BIS as the motivation in these streams is similar: how to use data effectively to impact business decision-making, thereby creating a competitive advantage. Additionally, given the blend of terms and conceptualizations of the phenomenon, it

is important to consider the streams as somewhat joined as technologies that can be leveraged to impact firm performance.

Chen et al. (2012) provided a unified term, Business Intelligence & Analytics (BI&A), thereby treating big data analytics as a related and connected field for research. Furthermore, Chen et al. (2012) noted that "Data analytics refers to the BI&A technologies that are grounded mostly in data mining and statistical analysis" (p. 1174). Considering the purpose of the current research, it is possible that SMB representatives may be more familiar or more experienced with BI&A, BA, or simply analytics terminology and concepts. In fact, some researchers whose stated intent was to study BDA have used these terms interchangeably in research data collection including survey items. Consequently, this research study will attempt to account for variation in definitions of data analytics and SMB representatives' understanding of the concept of DA and BDA in the study design and methodology. In a similar vein, it is important to understand any prior inquiries on SMBs that have considered or studied the use of these related technologies to realize business value.

Grover et al. (2018) noted the original three V's that are used to describe big data as volume, velocity, and variety:

> Volume refers to the ever-growing large magnitude of data. Velocity refers to the fact that data are generated and arriving continuously at an unparalleled speed and must be dealt with in a timely manner. Variety means that data are in diverse formats, ranging from structured data to unstructured data (e.g., text documents). (p. 389)

In addition, Grover et al. (2018) and Chiang et al. (2018) extended the conceptualization to add veracity, or accuracy and quality of data, and a key fifth dimension, value. As Chiang et al. (2018) described:

> In addition to the four V's (volume, velocity, variety, and veracity) characterization of big data, value has been considered the fifth key dimension in BDA. Analysis of data without generating value offers no contribution to an organization, regardless of whether data are big or small. The success of BDA projects requires not only infrastructure, data analysts, and knowledge and tools for dealing with big data, but also an *understanding of how BDA translates to competitive advantages and strategic value* [emphasis added]. (p. 348)

Based on their review of the literature, Mikalef et al. (2017) noted several additional V's, including variability or "how insight from media constantly changes as the same information is interpreted in a different way, or new feeds from other sources help to shape a different outcome" and visualization or "interpreting the patterns and trends that are present in the data" (p. 555). Interestingly, Mikalef et al. (2017) found based on their review of previous literature that "[w]hile there are no universal benchmarks for defining the volume, velocity, and variety of big data, *the defining limits are contingent upon size, sector, and location of the firm, and are subject to changing limits over time* [emphasis added] (Gandomi & Haider, 2015)" (p. 552). More importantly, what researchers have recognized is that the 'big' association can change over time. While historically associated with size (i.e., number of rows and columns), the lower cost of storage, increased processing power, emergence of sensors and connected devices, development of network infrastructures, and growth in cloud computing have made it easier to

generate reliable data analysis without significant expense (Mikalef et al., 2020a). Importantly, Grover et al. (2018) recognized the following:

> Today, a new, more pragmatic view of big data is taking hold, dominated not by discussions about the volume, velocity, and variety of data, but by the value of data—the ability to generate actionable insights and apply them to business practices to accelerate innovation, drive optimization, and improve business performance. (p. 400)

The important distinction from the earlier conceptions of DA is that numerous authors have recognized value as a key defining characteristic.

Therefore, business value creation is an important consideration for any organization that chooses to devote resources toward the development of analytics capabilities. For similar reasons, research focused on best practices associated with DA adoption and use may fall short in generalizability and potential impact on business practice if it does not incorporate considerations of realized business value. As Hazen et al. (2012) argued, “the mere adoption of an innovation by an organization does not necessarily imply that the innovation is actually being used or adding value to the firm and its trading partners” (p. 1).

With a focus on BDA, prior research has provided some evidence that DA has the potential to create value in large businesses likely due to the ability to conduct econometric analysis with an objective performance measure as the DV, or at least directly verify contributing or facilitating factors and their relationship with tangible performance measures (Brynjolfsson et al., 2011; Muller et al., 2018). However, it is important that the research and our understanding of the relationship between DA and firm performance and of the factors that lead to value creation from DA be extended to SMBs. Additionally, as Mikalef et al. (2020b) stated:

> Most reports to date on the business value of big data have been from consultancy firms, popular press, and individual case studies, which lack theoretical insight. There is, as a result, limited understanding on how firms should approach their big data initiatives and inadequate empirical support to support the claim that these investments result in any measurable business value. (p. 1)

In other words, it should not be left to SMB owner trial-and-error, consultancies, and software vendors to be the primary contributors to learning the 'what' and 'how' of leveraging DA to impact firm performance.

DABV Creation within SMBs

The objective of this dissertation is not to explain what factors lead to adoption and use of DA, but to identify capabilities that contribute to DA adoption within SMBs that results in the realization of business value. In other words, the purpose is not to explain the factors that lead to adoption but to identify the DA capabilities that contribute to DABV.

While SMBs may collect less data than large firms (i.e., less observations/rows, fewer variables, fewer sources), there are still opportunities for SMBs to leverage data to their benefit. Shah et al. (2017) argued many SMBs "collect, store and analyse higher amounts of transactional data in digital formats, and hence with the use of Big Data applications it can enable them to higher strategic and performance related information within their environments" (p. 4). However, the dominant focus by researchers and practitioners has been on successful generation of business value from data analytics at larger firms.

Thus, the third element of significance of this dissertation research can be described as business value, realization of benefits, or improvement in business performance. In other words, for those SMBs that have adopted and are in different stages of maturity, does the use of DA

result in different business value or performance outcomes and, if so, how does a firm achieve such outcomes? For example, Grover et al. (2018) argued, "…no single business trend in the past decade has had as much potential impact on incumbent IT investments as BDA" (p. 390). Thus, SMB leaders should be aware of how to go about developing capabilities that allow investment in data analytics resources to generate positive business performance outcomes. Grover et al. (2018) further elaborated that, "The ultimate success of any big data projects lies in realizing strategic business value, which gives firms a competitive advantage" (p. 390). With respect to how value is created, Grover et al. (2018) argued, "When examining successful cases of value realization, it is clear that BDA requires investments in data assets, technological assets, and human talent in order to generate something novel and valuable" (p. 397).

Accordingly, Grover et al. (2018) proposed a value creation framework based on resources of the firm derived from the resource-based view (RBV), also known as resource-based theory (RBT). Barney (1991) brought RBV to the forefront of management theory and the theory has since matured over time. More recently, RBV has been argued to be "one of the most prominent and powerful theories for describing, explaining, and predicting organizational relationships" (Barney et al., 2011, p. 1300). Consequently, RBV refers to the concept that resources and capabilities can explain the creation of a competitive advantage, where resources and capabilities are comprised of bundles of firm assets (Barney et al., 2011). With respect to BDA, Grover et al., (2018) proposed a conceptual framework for value creation which included BDA infrastructure (e.g., big data asset, analytics portfolio, human talent) leading to BDA capabilities (e.g., ability to integrate, disseminate, explore, and analyze data), leading to value creation mechanisms that lead to value targets (e.g., business process improvement, organization performance, products and services innovation, and consumer experience and market

enhancement) that then create impact (Grover at al., 2018). In the context of practice, Grover at al. (2018) argued that the challenge in developing analytical and data capabilities to generate value is multi-faceted and requires an alignment of resources:

> Importantly, the value realized from BDA is subject to weak links in the value creation process. For instance, excellent tools and data scientists will not help if the data are of poor quality, and big data experts may leave if a company is not committed to provide necessary resources to support analytics and deploy and benefit from the insights discovered from big data. (p. 419)

As a suggested direction for big data research, Grover et al. (2018) reported that in building analytics capabilities, one area for researchers with respect to analytics could be maturity focused or alternatively understanding the relationship of capabilities and adoption for realizing value. For example, "Questions around identification of analytical capabilities for people and processes and the adoption of IT tools to realize them. This would include organizational-level issues of conceptualizing BDA maturity" (Grover et al., 2018, p. 416). Additionally, Grover et al. (2020) argued that research should be grounded in prior IS literature and better suited for practice. Indeed, this is precisely what this research seeks to achieve. Finally, the focus on SMBs in the current study addresses the additional gap in the literature and more importantly a gap in practitioner knowledge about how to effectively use DA within SMBs to create business value. As such, while the focus of this dissertation is on DA capabilities and potential impact on performance, by studying DA with elements of BDA, I will provide a contribution to BDR given the understudied SMB setting.

Theoretical and Practical Implications of Studying SMB DA Practices

For example, Grover et al. (2020) argued that BDR researchers have prioritized "the practical value of the predictive accuracy of a model over generalizable theory" (pp. 273-274) or alternatively have focused on using new data sets or variables generated by a single organization or applied a new statistical technique that have significance in a local setting but cannot be generalized to a broader number of firms. They asserted the following:

> Generalizability of findings will become harder unless we establish our field's research criteria to include the expectation that produced knowledge needs to contribute to a cumulative tradition in which the aim is to understand behaviors and features of sociotechnical systems at a theoretical level. BDR, however, will likely underemphasize the importance of theory. (Grover et al., 2020, p. 277)

While research that involves mining data and performing some sophisticated modelling is enticing as it avoids the steps one must go through as part of completing the research such as Institutional Review Board (IRB), approval these studies can be more challenging to generalize for a practitioner audience. In context, if an SMB owner is seeking a prescription of management practices that should be adopted to use data effectively, analysis of an interesting industry data set is not that. For example, top IS researchers warned recently, "…many BDR papers seek to make strong claims with regard to their contribution based on the novelty of data (referring to the size or uniqueness of the dataset) or analytic technique used (e.g., application of a novel, rarely used before machine learning technique)" (Grover et al., 2020, p. 277) which are less focused on disseminating theoretical knowledge. While isolated predictive analytics and data mining industry projects and the resulting publications are useful for practitioners to understand what types of variables might be important and what techniques can be used, they often do not provide

guidance on the development of capabilities that lead to the explicit realization of business value. Thus, I seek to maintain the goal of IS research by building on previously established IS theory. Finally, I seek to understand a unique phenomenon with the hope that the findings and results can be generalized and the discoveries are more likely to influence actual practice in an understudied but important business setting.

Summary of Introductory Research Concepts

I have defined three aspects of DA that will be pursued in this research: (1) the importance of studying DA use within SMBs, (2) the resources and capabilities that influence the actual realization of DABV within SMBs, and (3) the potential business value than can be realized from DA within SMBs. While one may tackle these pursuits in different ways, my goal is to additionally contribute to and address the needs of the field of Big Data Research (BDR); however, it must be stated that the focus in this research setting is DA use within SMBs. I argue that this is achievable by focusing on generalizable research that addresses both an academic audience and business practitioners.

Purpose Statement: Research Objectives

The purpose of this dissertation is to understand DA practices within SMBs. More specifically, the purpose is to better understand how the maturity of DA resources and development of DA capabilities can impact SMB business performance. Consistent with the literature, I seek to identify resources that lead to analytics capabilities and may impact DABV. My motivation is to fill the gap in the literature pertaining to SMBs and to generate generalizable DA research based on reported SMB experiences while building on established theory. In addition, I seek to contribute to the establishment of effective DA practices for DA value realization in SMBs, thereby enhancing their competitiveness.

How Aims were Accomplished

I conducted an extensive literature review. Based on the literature review, I developed a conceptual and empirical model to test the relationships of between DA use, DA capabilities, and DABV. Drawing upon prior research, a survey instrument was designed that leveraged existing instruments that have been previously empirically validated by top IS researchers. As part of the research design, I obtained a sample of respondents from SMBs. The survey was administered to the sample and results analyzed to understand patterns and test and analyze the relationships between the factors in the postulated model.

Significance of the Dissertation

Consistent with the goal of influencing practice while contributing to IS research, I provide practitioners with a rigorous, theory-based, empirical analysis of the rarely studied phenomenon of effective DA use within SMBs and the necessary resources and capabilities associated with realization of DABV within SMBs. The research seeks to understand SMB owners' and management's perceptions, firm practices, and answers to two primary questions that many SMB owners, decision-makers, and private capital investors seek to understand:

- Does investment in DA impact business performance within SMBs?
    - More importantly, where should an SMB owner invest to establish an effective value-creating Data Analytics Capability?
- How does an SMB best manage use of DA based on the allocation of resources and development of data analytics capabilities to enhance the realization of business value?

CHAPTER 2: LITERATURE REVIEW

From a practical business perspective, the focus of this dissertation can be generalized to the effective use of DA to support decisions that positively impact business performance. For example, Brynjolfsson et al. (2011) stated that, "[in] more and more companies, managerial decisions rely less on a leader's "gut instinct" and instead on data-based analytics. At the same time, we have been witnessing a data revolution; firms gather extremely detailed data from and propagate knowledge to, their consumers, suppliers, alliance partners, and competitors" (p. 1). Based on interviews with executives at 330 public companies in North America and performance data collected, companies in the top one-third of their industry that leverage data-driven decision making (also known as Evidence-Based Decision Making or EBDM) are 5% more productive than competitors and 6% more profitable than competitors (McAfee & Brynjolfsson, 2012). Brynjolfsson et al. (2011) pointed out the following attributes of the business systems that collect data: "Increasingly these systems are imbued with *analytical capabilities, and these capabilities are further extended by Business Intelligence (BI) systems that enable a broader array of data analytic tools to be applied to operational data* [emphasis added]" (p. 1). Thus, there are underlying technical and human resources and processes that form capabilities that make EBDM possible. While this dissertation does not explicitly seek to study EBDM, I note that the broader context of EBDM explains the shift in managerial actions that might result from DA use that ultimately leads to the realization of business value.

DA technologies have advanced rapidly so there is a clear opportunity for firms to leverage these technologies in combination with other organizational capabilities such as human skills and redesigned decision-making processes. More importantly, the relevance of DA in business practice is that firms may develop analytical capabilities that can impact firm

performance outcomes (Gupta & George, 2016, p. 3). Finally, terminology and concepts surrounding DA have been opaque. Thus, it is imperative to understand the breadth of research that has emerged from different streams of literature that are focused on DA and recognize the universe of relevant theoretical models, identify capabilities that contribute to effective data analytics use, and develop appropriate measures of the business value that subsequently emerges.

The Development of DA Capabilities

Insights from the Literature on BI and BA

Puklavec et al. (2018) leveraged prior research on diffusion of innovations (DOI) and the technology, organization, and environment (TOE) framework to study BIS adoption stages within SMBs. Puklavec et al. (2018) noted that "adoption and use are two distinct stages. With BIS use we understand the stage in which BIS is widely used as an integral part in a firm's value chain activities and where its use helps attain the goals for using the system" (p. 4). Accordingly, for this dissertation, it is important to better understand the DA capabilities that may influence DA use and ultimately influence the realization of business value.

Kulkarni et al. (2017) studied the antecedents of the development of a BI capability. "BI systems focus less on reducing costs or increasing operational efficiency and more on *increasing managerial effectiveness and for building competitive advantages* [emphasis added]" (Kulkarni et al., 2017, p. 517). They discovered two organizational factors, user participation in the ongoing evolution of BI and analytical decision-making orientation, as potential mediating variables that help explain the relationship between top management championship and the development of a BI capability. While management buy-in or the notion that adoption and the development of a capability comes from the top seems to be a common theme in IS research, an analytical decision-making orientation is a factor that is often discussed but not directly studied.

Accordingly, Kulkarni et al. (2017) argued that their "analytical decision making orientation construct represents the encouragement that employees at all levels of the organization perceive to make decisions based on information and evidence and to support ideas, opinions, proposals, and so on with facts and figures wherever possible" (p. 526). They found support that having an analytical decision-making orientation mediates the relationship between top management championship and BI system capability based on a sample of 486 experienced managers. In fact, this analytical decision-making orientation concept seems very similar to the managerial skills and data-driven culture concepts studied by Gupta and George (2016) on the context of BDA.

In addition to BI, many prior studies of DA were focused on BA. Given the similarities of these areas and the crossover of concepts, definitions, success factors, and relationships that were examined empirically in various settings, prior research on BI and BA can apply to this study of DA. In terms of the focus of previous research, the explanation could be that there is a time element to emergence of different data analytics technologies. For example, some might argue that data or information that is generated from BA is BI (Wixom et al., 2013). Alternatively, it is possible that given reporting came first, often described as part of BI, that BI is what eventually led to BA or more sophisticated analytics. In any case, a current definition provides that BA is "the application of models directly to business data. Business analytics involve using DSS tools, especially models, in assisting decision makers" (Sharda et al., 2020, p. 771). Whereas BI is described as a conceptual framework for managerial decision support that "combines architecture, databases (or data warehouses), analytical tools, and applications" (Sharda et al., 2020, p. 771). Yet, "Indeed, many practitioners and academics now use the word analytics in place of BI. Although many authors and consultants have defined it slightly differently, one can

view analytics as the process of developing actionable decisions or recommendations for actions based on insights generated from historical data" (Sharda et al., 2020, pp. 30-32).

With respect to BA, Chen and Nath (2017) sought to study the relationships between managerial perceptions of IT, BA maturity, and success. "Achieving BA maturity involves an evolutionary process of developing competencies in areas including data, information technology, analytics expertise, management support, and strategic orientation" (Chen & Nath, 2017, p. 62). They argued that adoption maturity models allow for the assessment of organizational capabilities, processes, and resources associated with effective adoption. They further argued that a maturity model can assist a firm with evaluating their business analytics competency.

Chen and Nath (2017) used perceptual measures of firm-level IT impact and strategic role of IT. They evaluated the relationship between perceived firm-level IT impact and perceived strategic role of IT as factors that impact elements of BA maturity. In contrast, the objective in this dissertation is to understand how DA capabilities lead to variation in perceived business value. In surveying 177 Chinese business executives, Chen and Nath (2017) found that of the three maturity factors considered (BA integration and management support, process-level benefits of BA, and technology and DA capabilities) that smaller organizations had the lowest level of maturity and were lower in the acquisition of identified technology and DA capabilities. Similarly with respect to firm size, Chen and Nath (2017) argued that "these findings suggest that BA efforts are resource intensive and organizations with sufficient financial, human, technology and organizational resources are more likely to achieve a higher level of BA maturity and success than those lacking in such resources" (p. 73). Importantly, Chen and Nath (2017) found that process-level benefits of BA had a positive impact on firm's overall BA success and

argued that this finding is consistent with Tallon et al. (2000) “which advocates the contribution of IT to firm performance to be gauged through the value created by IT for individual business processes, which, in time, will aggregate to the firm level” (Chen & Nath, 2017, p. 74). Finally, they found that technology and data analytics capabilities do not have a significant direct effect on BA success yet is mediated by process-level benefits of analytics. Importantly, it can be argued that perceptual measures of DA on process-level outcomes are sufficient to assess firm performance impacts of DA.

Wixom et al. (2013) provide additional perspectives that offer a foundation for further research. They conducted a case study analysis of GUESS Inc. that examined the value derived from BA and identified several types of benefits: transactional benefits (i.e., less paper, time savings, fewer meetings, reduced headcount, faster cycle time, and convenience); informational benefits (factual decisions, real-time decisions, single version of truth, business pattern discovery, and more collaboration), and strategic benefits (i.e., speed to market, improved business understanding, and reputation). While findings seem anecdotal, at the very least it provides evidence that an organization may see benefits or realize value from DA in different business process areas.

While different theoretical models have been considered by authors in studying analytics capabilities, the most common that has been used to conceptualize the relationship between firm resources, analytics capabilities, and business value is the resource-based view. What follows next is a foundational review of the resource-based view followed by a review of the more developed “Big” Data Analytics literature specific to the development of analytics capabilities.

The Resource-Based View for Analytics Capabilities

According to the resource-based view, it is possible for firms to develop a competitive advantage or enhance performance through the bundling of strategic resources and/or capabilities (Gunasekaran et al., 2017). Bharadwaj (2000) studied the performance effects of IT from a resource-based perspective. According to Bharadwaj (2000), “A potential framework for augmenting the conceptual analysis of IT's effects on firm performance is the resource-based view (RBV) of the firm which links the performance of organizations to resources and skills that are firm-specific, rare, and difficult to imitate or substitute” (p. 170). Resources are basic units of analyses in which firms can create a competitive advantage by accumulating and developing resources to create organizational capabilities (Bharadwaj, 2000). Capabilities are the organization’s ability to assemble, integrate, and deploy these valued resources (Bharadwaj, 2000). In the context of this dissertation, it has been argued that resources represent the input while a capability is a firm’s ability to deploy these resources strategically (Mikalef et al., 2017).

There are three key IT-based resources including tangible resources that include physical IT infrastructure components, human resources that include technical and managerial skills, and intangible resources that include knowledge assets, customer orientation, and synergy (Bharadwaj, 2000). Further, there was a significant positive relationship between superior IT capability and superior firm performance. Numerous researchers who have studied analytics capabilities, and the relationship between these capabilities and performance, have relied upon resource-based theory for guidance, and specifically intangible, human, and tangible resources as the inputs for capabilities. Authors have also argued that the choice of post-technology adoption constructs depends on the research setting, research question, and research innovation (Hazen et al., 2012). In the current study, given my goal is to understand how the development of

capabilities impact business value, the resource-based view provides an appropriate theoretical foundation to identify constructs that lead to the development of an analytics capability. For example, Mikalef et al. (2017) argued, "A basic premise of RBT is that the capability-building process can only take place following acquisition of a resource; therefore, developing capabilities is dependent on, and confined under, the types of resources a firm decides to accumulate" (p. 559). They noted that the most common conceptualization of resources that has been relied upon and can be extended to analytics capabilities is provided by Grant (1991) in which resources can be divided into tangible resources (e.g. financial and physical), human skills or human resources (e.g. employees' knowledge and skills), and intangible resources (e.g. organizational culture and organizational learning).

While the main purpose of this dissertation is to understand the resources and management practices required for DA Capabilities, the BDA capabilities literature contains more direct examples where the resources required to form a DAC were empirically tested, often with firm performance as the target. Additionally, the resources and management practices identified from empirical BDAC research is directly transferable to BA. Furthermore, the focus should not be necessarily on the 'big' characteristic as the primary influence on resources and management practices but more so on resources specific to DA, regardless of size or variety. In other words, the main categories of tangible, intangible, and human resources established by IS scholars do not require manipulation but instead can be conceptualized around DA. Thus, it is appropriate to consider practices and resources developed around BDA as easily transferrable to simply DA.

Analytics Capabilities in BDA Literature

DA and other related concepts have been conceptualized in different ways by both researchers and practitioners. For example, Chen et al. (2012) argued that the terms big data and

BDA have been utilized to define data and techniques in applications that are large and complex and that "require advanced and unique data storage, management, analysis, and visualization techniques" (p. 1166). Grover et al. (2018) defined BDA as the "application of statistical, processing, and analytics techniques to big data for advancing business" (p. 390). Regardless of the definition, the key is that there is a simplified understanding that DA comprises technologies that can be used by firms to increase business value but that require the development of specialized capabilities to create and sustain such value.

For example, in the context of SMBs, a survey conducted on behalf of SAP in April 2012 of 154 C-suite global executives revealed some respondents focused on "what [*Big Data*] is, while others tried to answer what [*Big Data*] does" (Gandomi & Haider, 2015, p. 132). Yet, it has been recognized that value is achievable by SMBs as these firms can use BDA to analyze voluminous, semi-structured data to positively impact firm processes (Gandomi & Haider, 2015). Thus, how DA and its potential to generate value are understood by SMB firm respondents can and should be addressed in the research design. In other words, I recognized the possible need to adjust research activities based on perceptions and understanding of DA concepts in a new and understudied setting. In any case, this leads to the following research question and hypothesis:

RQ1: Does Data Analytics create value for SMBs?

H1: DA creates business value within SMBs through positive performance impacts on key business process areas.

Moving away from definitions around big data, the literature around analytics capabilities can now be discussed further. Gupta and George (2016) argued that "[w]hile in the 1980s, IT was touted as a competitive weapon, currently, it is big data that is heralded as the next big thing for organizations to gain the competitive edge" (p. 1). The question then becomes, what explains

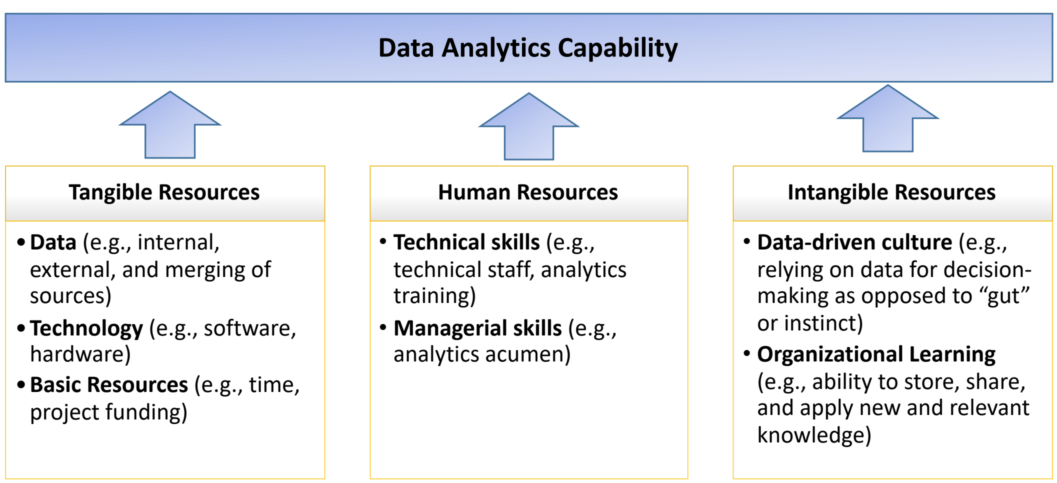
how a firm goes from investing and adopting to reaping benefits from data analytics? Alternatively, this can be distilled into the following:

- Which resources are most important for realizing value from DA?
- What leads to the development of capabilities around DA?

Gupta and George (2016) argued that to create a capability a firm needs an inimitable mix or combination of its financial, physical, human, and organizational resources and needs to continually adapt their resources based on changes in market conditions. More importantly, they identified seven different resources required to build an analytics capability. These include "data, technology, and other basic resources (e.g., time and investments)" which they describe as tangible resources, "managerial and technical big data skills" which they describe as human resources, and "data-driven culture and the intensity of organizational learning" which they describe as intangible resources (Gupta & George, 2016, p. 3). Figure 1 shows an adapted summary of the classification of resources that lead to an analytics capability.

Figure 1

Classification of Data Analytics Resources



Note. Adapted from Gupta and George (2016)

Gupta and George (2016) gathered survey data from 108 executive-level technology leaders and found that the results of the study “empirically validated the relationship between BDA capability and firm performance” (p. 13).

Grover et al. (2018) argued that to develop BDA capabilities that can impact business value, investment is required in infrastructure which includes big data asset, analytics portfolio, and human talent. Big data infrastructure includes data sources and platforms required to collect, store, process, and analyze data. The analytics portfolio includes exploring different ways to integrate and analyze data for different purposes (i.e., consumer sentiment analysis, financial risk modeling, marketing campaign analysis, cross-selling, fraud detection, recommendation improvement, and price and performance optimization). Human talent includes proficiency and experience with DA (Grover et al., 2018). Additionally, there are six distinct mediating mechanisms that link capabilities and value targets: transparency and access, discovery and experimentation, prediction and optimization mechanisms, customization of products and services, learning and crowdsourcing, and ability to monitor situations and adapt rapidly (Grover et al., 2018). Thus, what is suggested is that the presence of capabilities can lead to different business value outcomes. Another key point is that moderators or ‘contextual enablers’ could include a prevailing data-driven culture that requires “alignment of a strategy with the organizational infrastructure” that in turn requires strong leadership (Grover et al., 2018, p. 404).

Other authors have also considered the impact of capabilities on firm performance in the context of DA. For example, Wamba et al. (2017) studied the link between capabilities and firm performance. They noted the challenges of evaluating the link between IS investments and firm performance as including the following: unavailability of data, time lags between investments and business value generated, lack of proposer assessment of the indirect benefits of IS, and

deficiencies in the level of analysis of IS-related benefits. They studied the links between management capabilities, infrastructure capabilities, and personnel capabilities based on the resource-based view and socio-materialism theory. In surveying 297 Chinese IT managers and business analysts, Wamba et al. (2017) studied BDA management capabilities by referring to the concept of BDA as business analytics. Based on the survey design, they essentially argued that when collecting information on BDA there seems to be varying ways to reference the phenomena, likely so that respondents can better conceptualize the items in the questionnaire. Particularly, within an SMB setting where level of maturity or knowledge may be behind that of larger firms, it appears that previous research has justified framing the terminology to make it more digestible for research participants and practice. Wamba et al. (2017) found a positive relationship between both BDA capabilities and process-oriented dynamic capabilities (PODC). More specifically, these variables explained 65% of the variance of firm performance, measured using perceptual measures, while 30% of the variance was explained by the mediator, PODC. Thus, BDA capabilities influence firm performance and set a precedent for adjusting terminology for the empirical setting. However, their study may have some generalization limitations as their sample consisted of IT managers and business analysts at larger firms and was limited to China.

Gunasekaran et al. (2017) studied the relationship between big data and predictive analytics (BDPA) and firm performance. Gunasekaran et al. (2017) argued:

> RBV is relevant for understanding BDPA assimilation as a capability that is dependent on bundling C [Connectivity] and IS [Information Sharing] (resources), and impacts positively on SCP [Supply chain performance] and OP [Organizational Performance] and subsequently to the achievement of competitive advantage… (p. 314).

Connectivity refers to the information systems ability to link and communicate with firm partners where information sharing is the quality of the information shared (Gunasekaran et al., 2017). Gunasekaran et al. (2017) define assimilation as the organization's ability to diffuse BDPA across organizational processes. In addition, top management commitment mediates the relationship between information sharing and connectivity and BDPA acceptance. Further, an organization should improve BDPA acceptance and assimilation capabilities via the mediating construct referred to as BDPA routinization or how a firm's governance systems are adjusted effectively to facilitate BDPA. Based on a survey of 205 "key informants who are functional heads associated with SCM (logistics/transportation, operations management, and purchasing/procurement)" in India, BDPA assimilation enhances performance (Gunasekaran et al., 2017, p. 312).

The contribution of this research and previously mentioned DA research is that it has been shown that resources focused on DA, when bundled, can lead to the development of a capability that can impact business value. However, this dissertation seeks to identify the key resources required to develop data analytics capabilities that are both empirically validated and can be easily understood by SMB practitioners, which is more characteristic of intangible, human, and tangible resources (Gupta & George, 2016).

Mikalef et al. (2017) conducted a systematic literature review of BDA capabilities founded on the resource-based view of the firm and dynamic capabilities view (DCV). Mikalef et al. (2017) noted that "[t]he review process was driven by the following research question: What are the definitional aspects, unique characteristics, challenges, organizational transformations, and business value associated with big data?" (p. 549). They argued that big data, BDA, and BDA capability are often used interchangeably in the literature and that to adapt

to the data-driven era researchers have begun to separately define BDA capability as “a company’s proficiency in leveraging big data to gain strategic and operational insight” (p. 552). Importantly, a BDAC does not refer to use of big data alone but should include organizational resources that are important in leveraging data for strategic gain (Mikalef et al., 2017). Mikalef et al. (2017) further borrowed from Gupta and George (2016), noting three main categories of resources to develop a BDAC are tangible resources, intangible resources, and human skills and knowledge resources. Mikalef et al. (2017) argued that industry and contextual factors may impact firms’ decisions to employ analytics capabilities. Additionally, how firms choose to leverage analytics capabilities could differ and as a result lead to differences in performance gains. To this point, “It is therefore important to construct specific performance measures depending on a number of contextual factors, as well as on the area in which BDA are deployed” (Mikalef et al., 2017, pp. 571-572). Thus, contextual circumstances matter and firm strategy and focus area could impact analytics capability outcomes. “As such, a firm can have developed a strong BDA capability but only utilize it towards a specific type of operational capability (e.g., marketing)” (Mikalef et al., 2017, p. 570). Consequently, this dissertation will attempt to understand differences in DA capabilities in relation to strategic priorities for SMB firms and how these capabilities lead to value realized within different firm process areas.

Mikalef et al. (2019) utilized a mixed-method approach to study the relationship between DA and firm performance in Greek firms. They relied on input constructs that included data, process, technology, organization, people, and context with the dependent variable being firm performance. Mikalef et al. (2019) argued that “high levels of maturity in these specific elements enable the firm to achieve performance gains under a certain set of contextual conditions” (p. 263). In other words, through a complexity theory lens and configurational methodologies, it is

possible that there are different combinations of inputs from which firms might achieve high levels of performance from BDA (Mikalef et al., 2019). They concluded that there is still limited research on how the importance of certain resources may differ based on the context and in turn how a mix of factors can influence performance depending on the setting. Yet, their study borrows heavily from previous research. While largely relying on the constructs developed by Gupta and George (2016), Mikalef et al. (2019) included several contextual variables and argued that structural and relational organizational factors can help explain different performance outcomes. Specifically, "[s]tructural practices include explicit declarations about the main roles of setting policies and standards for protecting and using data. Relational practices on the other hand are concerned with the formalized links between employees of the technical and business sides" (Mikalef et al., 2019, p. 266). In addition, Mikalef et al. (2019) included processes:

> Processes regarding big data analytics refer to the formal methods for managing and leveraging data in order to generate insight. In this regard, Procedural (PROC) practices – typically part of big data governance – are concerned with activities that amongst others include data migration, data retention, cost allocation, data analytic procedures, and access rights (p. 266).

Based on 175 responses from IT managers in Greece, Mikalef et al. (2019) found, using fsQCA, four alternative solutions that lead to high performance including two for large firms and two for SMBs. This demonstrates that SMBs may have different DA-related levers that lead to realization of business value, thus motivating the need to study these firms separately from large firms. Furthermore, it demonstrates that several factors need to be considered when investing in DA to understand which allocated resources can be bundled to impact firm performance. However, a main goal of this dissertation is to provide clarity for practitioners; therefore,

providing some common factors that can be identified and assuming a more linear relationship between factors and business value, the methodology and modelling approach employed by Gupta and George (2016) seems more appropriate.

Mikalef et al. (2020b) studied the relationship between BDAC and competitive performance, investigating the mediating roles of dynamic and operational capabilities. They surveyed 202 CIOs and IT managers in Norwegian firms and found that BDAC can help develop competitive advantage. The model tested the relationship between BDAC and dynamic capabilities, where dynamic capabilities was examined as a mediator between BDAC and Marketing Capabilities (MC) and Technological Capabilities (TC). Mikalef et al. (2020b) noted:

> [a]s the direct effect of BDAC on MC is found to be nonsignificant and the mediating path is found to be significant, we can conclude that dynamic capabilities fully mediate the effect of BDAC on marketing capabilities. On the other hand, as the direct effect of BDAC on TC is still significant and the mediating path is also significant, we demonstrate that dynamic capabilities partially mediate the effect of BDAC on technological capabilities (p. 11).

Based on research focused on the role of information governance in BDA, Mikalef et al. (2020c) found “a firm’s BDAC helps enhance two distinct types of innovative capabilities, incremental and radical capabilities, and…information governance positively moderates this relationship” (p. 1). Mikalef et al. (2020c) define information governance as a collection of competences or business practices that focus on the “creation, capture, valuation, storage, usage, control, access, archival, and the deletion of information and related resources over its life cycle” (p. 3).

Consistent with RBV and its application in BDAC research, Mikalef et al. (2020c) noted, “Grant proposed a distinction between tangible resources (financial and physical), human skills

(employees' knowledge and skills), or intangible resources (learning propensity and organizational culture)" (p. 2). This means that information governance has a role in enhancing business benefits: "information governance will only have positive effects if it is complemented with the presence of a strong BDAC" (Mikalef et al., 2020c, p. 5). Thus, enhancing the value derived from analytics capabilities requires information governance, jointly comprised of structural, procedural, and relational practices. Based on a sample of 175 CIOs and IT managers in Greece, Mikalef et al. (2020c) found a positive and significant effect from BDAC on incremental innovative capability (INC) and radical innovative capability (RAD), with the moderator, information governance, having a positive and significant effect for the relationship between BDAC and RAD

This dissertation follows Gupta and George's (2016) suggestion that the emphasis should be on building analytics capabilities as opposed to explicitly studying data characteristics:

> We do not suggest that discussing (big) data characteristics is not important; however, the over emphasis on big data characteristics steers focus away from the critical question that the organization must be asking: How to create big data capabilities, which in turn may lead to superior firm performance? (pp. 2-3).

Indeed, data is a resource that contributes to the formation of an analytics capability when combined with other resources such as technical skills to analyze it and management buy-in to rely on such analyses when making key decisions. Additionally, due to the limited empirical research on analytics capabilities specific to SMBs, I reviewed and considered the global literature on DA capabilities, much of it described as BDA, that has mixed the study of both SMBs and large corporations. This leads to the following research question:

RQ2: What is the importance of resources and organizational capabilities that contribute to the development of a DAC within SMB firms?

H2: Tangible, human, and intangible resources are equally important in the development of a DAC.

A summary of the key research performed surrounding capabilities is shown in Table 1.

Table 1

Summary of Analytical Capabilities Research Studies

Research Field	Author(s)	Journal	Theoretical Model(s)	First Order	Second Order	Capabilities	Primary Dependent Variable(s)	Key Findings
Business Intelligence	Kulkarni, Robles-Flores, & Popovic, (2017)	Journal of the Association for Information Systems	Structurational Model of Technology	Top Management Championship -> User Participation and Analytical Decision-Making Orientation		Information Capability and Business Intelligence System Capability	Information Capability and Business Intelligence System Capability	Support that Analytics Decision-Making Orientation mediates the relationship between Top Management Championship and BI System Capability
Business Analytics	Chen & Nath (2017)	Information Systems Management	Existing Maturity Models; Sociotechnical system theory	Perceived Firm-Level IT Impact and Perceived Strategic Role	BA Maturity Factors: BA Integration and Management Support, Process-Level Benefits of BA, and Technology & Data Analytics Capabilities		BA Success	Support that process-level benefits of BA have a positive impact on BA success; technology and data analytics capabilities do not have a significant direct effect on BA success but instead is mediated by process-level benefits of analytics
Big Data Analytics	Gupta & George (2016)	Information & Management	RBV	Data, Technology, Basic Resources, Managerial Skills, Technical Skills, Data-driven Culture, Intensity of Organizational Learning	Tangible, Human, and Intangible	BDA	Market Performance and Operational Performance	Support that BDA Capability has a positive impact on firm performance
Big Data Analytics	Wamba, Gunasekaran, Akter, Ren, Dubey, & Childe (2017)	Journal of Business Research	RBV; Sociomaterialism Theory	Connectivity, Compatibility, Modularity, Planning, Investment, Coordination, Control, Technical Knowledge, Technology Management Capability, Business Knowledge, Relational Knowledge	BDA Infrastructure Flexibility, Management Capabilities, and Personnel Expertise Capability	BDA Business Analytics (Third-Order) and Process-Oriented Capabilities (First-Order)	Firm Performance: Financial Performance and Market Performance	Support that there is a positive relationship between Big Data Analytics Capabilities and Process-oriented Dynamic Capabilities; both have a positive relationship with firm performance
Big Data and Predictive Analytics	Gunasekaran, Papadopoulos, Dubey, Wamba, Childe, Hazen, & Akter (2017)	Journal of Business Research	RBV	Connectivity, Information Sharing -> Top Management Commitment -> BDPA Acceptance -> BDPA Routinization -> BDPA Assimilation			Organizational Performance and Supply Chain Performance	Support that BDPA Assimilation enhances performance
Big Data Analytics	Mikalef, Boura, Lekakos, & Krogstie (2019)	Journal of Business Research	RBV; Complexity Theory	Data, Technology, Managerial Skills, Technical Skills, Structural Practices, Relational Practices, Data-driven Culture, Procedural Practices, Dynamism, Heterogeneity, Hostility Note: The authors used fsQCA to identify solutions that lead to high performance.			Performance	Four alternative solutions in which BDA practices and other factors can lead to high performance including two for large firms and two for SMBs
Big Data Analytics	Mikalef, Krogstie, Pappas, & Pavlou (2020b)	Information & Management	RBV; Dynamic Capabilities View	Data, Technology, Basic Resources, Managerial Skills, Technical Skills, Data-driven Culture, Organizational Learning	Tangible, Human, and Intangible	BDA	Dynamic Capabilities -> Marketing Capabilities, Dynamic Capabilities -> Technological Capabilities, Marketing Capabilities -> Competitive Performance, Technological Capabilities -> Competitive Performance	Support that dynamic capabilities fully mediate the effect of BDAC on Marketing Capabilities; dynamic capabilities partially mediate the effect of BDAC on Technological Capabilities
Big Data Analytics	Mikalef, Boura, Lekakos, & Krogstie (2020c)	Information & Management	RBV	Data, Technology, Basic Resources, Managerial Skills, Technical Skills, Data-driven Culture, Organizational Learning, Structural Governance, Procedural Governance, Relational Governance, Dynamism, Heterogeneity, Hostility	BDA: Tangible, Human, and Intangible; Information Governance; Environmental Uncertainty	BDA	Incremental Innovative Capability and Radical Innovative Capability	Support that BDAC has a positive effect on Incremental Innovative Capability and Radical Innovative Capability; the moderator, Information Governance, has a positive and significant effect for the relationship between BDAC and Radical Innovative Capability

Data Analytics Business Value (DABV) and Impacts on Firm Performance

Data has been considered a valuable asset, yet it is rarely shown as such in financial reports. For example, Tallon (2013) argued,

> When information assets are omitted from the balance sheet, they are treated as having zero value even in the most data-rich organizations. In the absence of an express or justifiable requirement for valuing data, organizations have failed to create methodologies that accurately measure the value of information or that track changes in value over time (p. 27).

This dissertation is focused on how data analytics capabilities (i.e., the capabilities that can process raw data into useful information and insights) can impact firm performance. Grover et al. (2018) argued that DA value is revealed through the combination of insight generation and its actual use. In fact, Grover et al. (2018) went on to state, "Academics have not studied how BDA can lead to a competitive advantage and strategic business value" (p. 391). They argued that realized strategic business value can be both functional (e.g., market share, financial) performance based and symbolic (e.g., positive brand image and reputation, mitigating environmental pressure). Additionally, Grover et al. (2018, p. 402) defined BDA value targets as:

- Organization performance such as the quality of decision making
- Business process improvement
- Product and service innovation
- Customer experience and market enhancement

Interestingly, these BDA value targets essentially focus on specific processes and function areas where firms can leverage DA to impact firm performance.

Given the business challenges with collecting direct measures of performance resulting from IT investments (i.e., actual change in revenue dollars, actual change in net income dollars, actual reduced costs, etc.), this dissertation adopts a process-oriented approach to assessing business value first proposed by Mooney (1996) and Mooney et al. (1996) that was subsequently further developed by Tallon et al. (2000) and other scholars. There has been some criticism regarding whether perceptual measures of business value are suitable proxies for actual impacts on firm performance. However, scholars have provided evidence to counter such claims. Tallon and Kraemer (2007) utilized a survey of executives representing 196 firms and found that executive "perceptions of IT impacts at both the process and firm levels are sufficiently accurate, credible, and unbiased as to constitute a viable approach to IT impact assessment" (p. 45).

Tallon (2007) studied how business strategy and IT strategy impact strategic alignment and how firm strategic alignment impacts business value. To capture business value, IS perceptual measures have been designed to capture managerial insight or intuition regarding IT impacts within and on various areas of the business (Tallon, 2007). Tallon (2007) argued that firms follow three primary strategic value disciplines: operational excellence, customer intimacy, and product leadership. These, in turn, suggest five process or functional areas that can be impacted by IT: supplier relations, production and operations, product and service enhancement, sales and marketing support, and customer relations (Tallon, 2007). Tallon (2007) requested respondents to identify the relative emphasis placed on the three strategic disciplines so that weight could be assigned to the business value contributions to their underlying processes:

> To identify this weighting, IT executives were asked to allocate 100 points across all three value disciplines with their primary value discipline receiving the highest number of points. Past research has found this to be an effective way of gaining

> insight into a firm's business strategy, while evidence of interrater reliability among executives in the same firm points to a shared understanding of strategy (p. 234).

This allowed for classification of firms as operationally excellent (OE), customer intimate (CI), product leaders (PL), or mixed foci based on whether the firm provided 50 or more points to a value discipline (Tallon, 2007).

I followed a similar approach to studying the business value of effective use of DA. However, as opposed to applying weights to values, business value was operationalized as a business performance outcome within the five process areas. I used perceptual measures of business value impact on the five process areas as a proxy indicator of the impacts of DA on SMB firm performance. Additionally, the value discipline grouping proposed by Tallon (2007) was used to test if there were differences in the relationship between DAC and business value across groups.

While Gupta and George (2016) sought to capture operating performance and market performance that they argued are "two important and distinct dimensions of firm performance in the IS literature" (p. 11), this dissertation focuses on how a DAC impacts process-level performance to account for variation in SMB firm strategy, industry positioning, product offerings, and process designs. The use of alternative performance/outcome measures is supported by other researchers who have studied the relationship between DA capabilities and value. For example, Mikalef et al. (2017) stated "we highly encourage researchers to examine a multitude of not only objective, but also subjective, value measures" (p. 572). This provides for the following research question:

RQ3: What is the effect of DAC on the realization of DABV within SMBs?

H3: DAC positively effects DABV as measured by perceived value in business process areas

The Role of IG

Another dimension that has been identified, information governance, has little literature and operationalization of the construct. For example, Mikalef et al. (2017) initially reported intangible resources as including information governance. Yet, Mikalef et al. (2020c) conceptualized information governance as a construct that is outside of DAC and expected to moderate the relationship between DAC and innovative capabilities. Mikalef et al. (2020c) argued, “structured adoption of big data analytics in the form of BDACs will lead to enhanced incremental and radical innovation capabilities, which will be amplified under the presence of information governance practices” (p. 2). In this sense, information governance represents a combination of “competences or practices for the creation, capture, valuation, storage, usage, control, access, archival, and the deletion of information and related resources over its life cycle” (Mikalef et al., 2020c, p. 3).

Mikalef et al. (2019) first introduced information governance in the context of empirical data analytics capabilities research, described as including structural practices, relational practices, and procedural practices. Additionally, Tallon (2013) suggested that possessing data governance practices help firms manage the value, risk, and cost of data. Yet, the investigation of what role information governance serves seems somewhat immature from a research perspective, at least when considering the relationship between the RBV-based DAC and performance measures. There is limited empirical evidence on the effects and mechanisms of information governance in an organizational setting (Mikalef et al., 2020c). Given that Mikalef et al. (2020c) treated information governance as a moderator between BDAC and innovative capabilities, it is

possible that it may similarly serve as a moderator between DAC and business value. Thus, it may be important to capture this factor to conduct a secondary test in addition to considering the baseline model.

According to Tallon (2013), information or data governance includes structural practices (e.g., user involvement in policy setting and evaluation, identifying responsible parties for data ownership, analysis of data value, and cost management), operational or procedural practices (e.g., actual organizational activities related to governance including enactment of retention policies, data migration, setting and monitoring of user access privileges, implementation of backup processes, establishing chargeback procedures to recover operating costs, execution of storage procedures, and establishing e-discovery and archival procedures), and relational practices (e.g., educating data users and non-IT managers on storage and costs and flow of communication across executives, IT, and data users). Mikalef et al. (2020c) argued, "The main responsibility of information governance is to answer the question: what information do we need, and how do we make use of it and who is responsible for it?" (p. 5). Accordingly, information governance can be conceptualized as a framework to optimize the value generated from firm data and information (Mikalef et al., 2020c). This allows for the final research question:

RQ4: What is the role of information governance in realizing DABV from DAC?

H4: Information Governance positively moderates the relationship between DAC and DABV or increases the effect of DAC.

Summary of Insights from the Literature

The resources and corresponding constructs identified by Gupta and George (2016) and leveraged by subsequent authors are key constructs for the development of an analytics capability. The key underlying theory that drove the identification of such constructs is based on

the resource-based view of the firm which has been accepted by IS scholars as a way of conceptualizing the business requirements for a capability. Further, it has been established that the DAC construct can have an impact on business performance operationalized by previous authors with various perceptual, business performance constructs. Additionally, process-driven business value measures that have been used to evaluate the impact of IT in previous research, similarly with the notion that IT requires investment in resources, may provide rich insights in the context of DA as they have been yet to be tested in such a setting. Information governance may also contribute to the success of DAC.

Next, I provide a synthesis of the takeaways from the literature to develop the main research model. According to Gupta and George (2016): "A firm needs a unique blend of its financial, physical, human, and organizational resources to create a capability, which will be difficult to match by competitors" (p. 1). Thus, the key resources considered in the development of a DAC to generate value or impact performance include: tangible (i.e., infrastructure), human (i.e., skill), and intangible (i.e., culture). Gupta and George (2016) argued, "Given that the main objective of this study is to identify several resources that will allow organizations to create BDA capabilities, which in turn may lead to superior firm performance, the choice of RBT as a theoretical framework for this study seems appropriate" (p. 2). Mikalef et al., 2020a) aggregated big data and BA as a phenomenon for enhancing business value. Mikalef et al. (2020a) found that "While there is significant variation with regards to the term used to denote this capacity, there is overall consensus about the key resources needed to develop a *big data analytics or business analytics capability* [emphasis added]" (p. 1). Additionally in reference to Gupta and George (2016) and Wamba et al. (2017),

> Recent studies in the area of big data and business analytics acknowledge that firms must develop a firm-wide capability of leveraging their big data and business analytics resources in order to realize value…This perspective has gained traction within the academic and practitioner communities as it sees big data and business analytics as organizational-wide investments that necessitate investments in resources at different levels throughout the organization (Mikalef et al., p. 7).

As a final point of justification for the RBV-based approach that has been applied, this dissertation takes the position of the relationship between DAC and competitive performance: "…building on the RBV and on prior studies on big data analytics, we define the notion of BDAC as the ability of a firm to effectively deploy technology and talent to capture, store and analyze data, toward the generation of insight" (Mikalef et al., 2020b, p. 2).

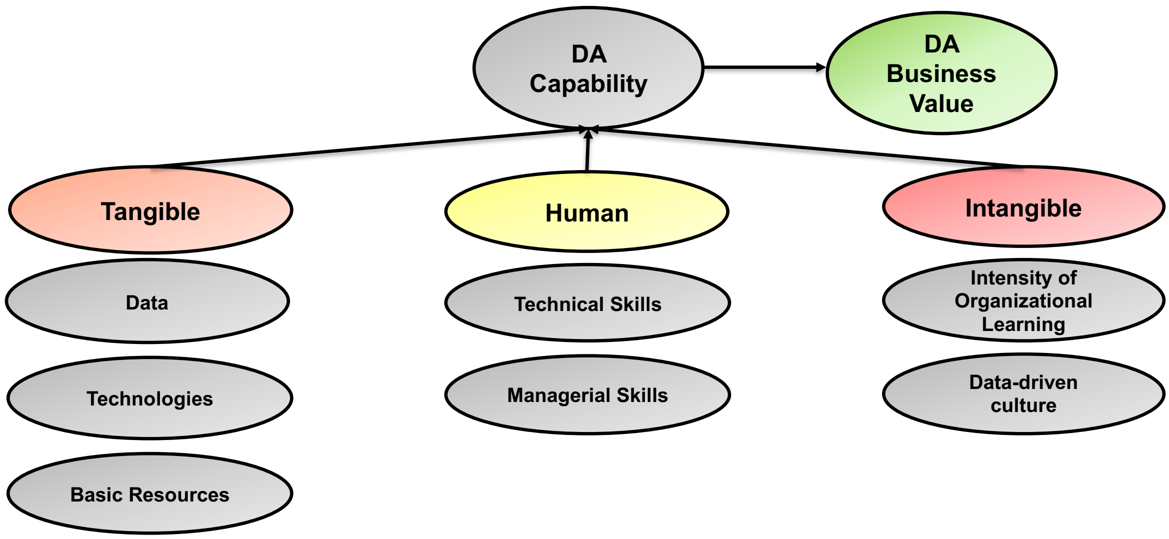
I intend to leverage the previously applied, RBV-based theoretical model presented to capture and understand the elements and maturity of DAC within SMBs and to test the potential impact on business value realized. Finally, I will use the process-driven business value measures as business value can be achieved in different strategy areas and thus richer insights can be derived around different performance areas than single measures of performance such as market performance.

Conceptual DAC Model

The research shows that there is consensus about the key resources needed to develop a BDA or BA capability (Mikalef et al., 2020a). It is then valuable to show the theorized DAC research model and required resources or constructs, shown as Figure 2.

Figure 2

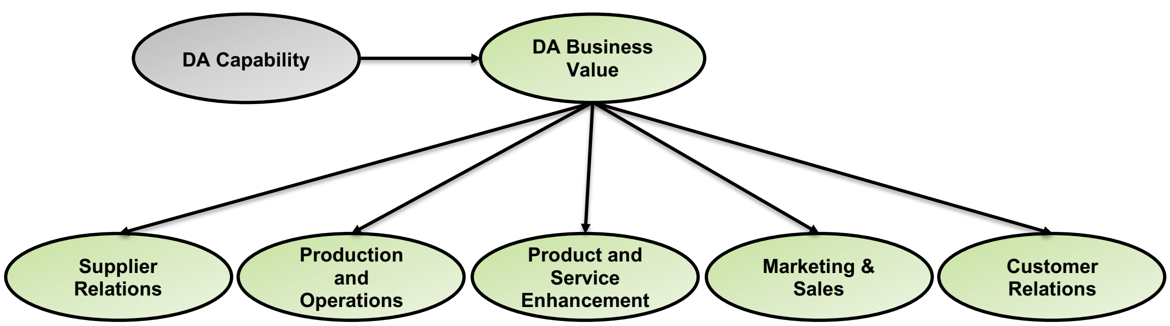
DAC



Furthermore, as shown in Figure 3, DABV will be captured by adapting perceptual measures of performance impacts at the process level developed by Tallon (2007).

Figure 3

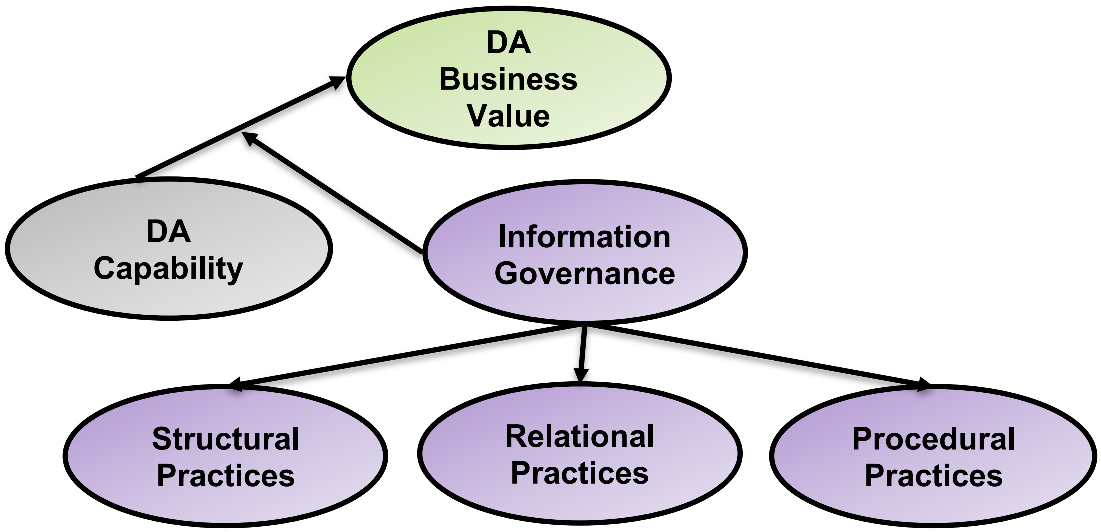
DABV



Information governance would be expected to moderate the relationship between DAC and business value. Yet, I see this as a secondary analysis to the main model. Based on the limited evidence available on the role of information governance in the context of the DAC research model, the theorized relationship is shown in Figure 4.

Figure 4

DA Information Governance



Mikalef et al. (2019) argued that it is important to understand contextual factors:

> While data, technology, people, processes and the organization of these comprise core components of realizing performance gains, the ways in which they are structured and the extent to which they are important is argued to differ depending on a number of contextual factors (p. 263).

In this dissertation, I propose that contextual factors do not in themselves necessarily lead to the formation of a DAC but may allow one to test for differences. It then may be more suitable to consider size and/or industry to evaluate differences across groups.

CHAPTER 3: RESEARCH DESIGN AND METHODS

Overview

The literature review highlighted that core concepts examined in this research have a robust theoretical base and have been examined in other business and organizational contexts. Therefore, based on the maturity of the research, a mixed-methods approach that includes both a quantitative study and qualitative research is appropriate.

Study Design

To investigate the dynamics of the relationship between DAC and SMB firm performance, an online survey was administered to a sample of SMB owners to capture data about the dynamics of resources and processes required for the development of a DAC, constructs for DABV, constructs for information governance, and other characteristics specific to SMB respondents. The survey was followed by a set of short case study interviews focused on helping me interpret the results and better understand the dynamics between key constructs. This mixed-methods approach is appropriate given the maturity of the underlying constructs and its ability to add additional insight beyond the information gathered using quantitative or qualitative methods alone (Creswell & Creswell, 2018).

The survey was developed by relying on previously utilized instruments based on research identified in the literature review:

- DA Capability (Gupta & George, 2016; Mikalef et al., 2019)
- Business Value (Tallon, 2013)
- Information Governance (Mikalef et al., 2019; Tallon et al., 2013)

The survey included a pre-test/pilot phase in which several respondents were interviewed after their responses and the data analyzed to obtain feedback. The survey was edited to incorporate

respondent feedback and findings from preliminary statistical analysis. After administering the final survey, five follow-up qualitative interviews were conducted to obtain additional context on the constructs that lead to the development of a DAC and how SMBs have been able to realize value from DA. These case study interviews were conducted with intentionally selected business owners to obtain additional insights around how SMBs can realize value from data analytics considering the main resources categories: tangible, human skills, and intangible. While the semi-structured interview guidelines were mostly original, insights and questions reported in Mikalef et al. (2019) provided the foundation for the development of the guidelines as this was the only example with a mixed-methods approach surrounding the identified DA research model. More specifically, after obtaining additional understanding from the quantitative analysis of the survey data, questions were edited and streamlined to capture more direct insights.

Study Population and Sampling

The target population was U.S. based, privately held SMBs. The sample was drawn from previous participants in the Private Capital Markets (PCM) project survey at Pepperdine University's Graziadio Business School. The PCM project has accumulated over 6,700 business owner contacts across the U.S. While the project has historically sought to obtain information from different groups, including angel investors and venture capitalists, the key group of interest related to this study was business owners. To supplement the responses received from the PCM contact list, I reached out to my personal network to identify additional SMB representatives. While response rates have historically been a challenge in IS research, I felt that this could be exacerbated during the COVID-19 environment particularly given that several businesses that had to either temporarily suspend operations or shut down permanently as result of the pandemic. Thus, because the response rate could have been impacted by what one could only

describe as a black swan event and to ensure the findings from the research were robust, I felt it was appropriate to supplement the PCM sample.

Focusing on the main mechanism of recruitment, in 2019 it was noted that 560 privately held businesses responded to the PCM survey. The firms represented ranged in industry and geography. In fact, many of the respondents were already accustomed to being contacted for participation in an annual survey. Those who participated in this study had previous experience in responding to questions related to investment activities, management practices, and firm performance changes across a specific time period. Further, the existing contact list was highly populated with qualifying U.S. SMBs based on size.

For the purpose of this research, firms with fewer than 500 employees were categorized as SMBs. While the definition based on employee count can vary depending upon industry and size can also be derivative of revenue or income, employee count seems to be a generally accepted standard for current business practitioners to study and analyze SMBs.

Given that the majority of the respondents were sourced from the same data set, approximately 80% of respondents had 500 employees or less. Additionally, while 65% of these respondents were controlling owners who actively operated the business, other types of managers also provided responses to the survey. Thus, when filtered to SMBs, this left almost 450 respondents. The 2019 PCM report indicated that approximately 41% of businesses had between one and five employees. Thus, I believed it was necessary to contact SMB representatives outside of the PCM contact list to provide some additional diversity in firm size representation. Based on the number of constructs and relationships in the model, the final sample exceeded minimum sample size requirements. Finally, recruitment via contact lists,

LinkedIn, and other digital forums appears to be a common, accepted practice in IS research (Mikalef et al., 2019, 2020b, 2020c).

The target population in this setting differs from previous studies which were focused on larger, global firms where the respondent often had a specific executive role or experience specific to information technology. In this dissertation, I sought respondents who had direct ownership of the business or a high level of operating control and knowledge about firm resources, processes, and performance.

Statistical power is defined as "the ability to correctly reject a false null hypothesis (in other words, to detect effects when indeed effects are present) and is calculated based on a particular effect size, alpha level, and sample size" (Osborne, 2013, p. 21). As Osborne (2013) summarized, "no prudent researcher would conduct research without first making a priori analyses to determine the optimal sample size to maximize the probability of correctly rejecting the null hypothesis" (p. 22). I attempted to ensure that the sample size was large enough to test hypotheses and sufficient in size as to avoid exhausting unnecessary resources to collect a larger sample.

Given that the approach to analyze the data was based on Partial Least Squares-Structural Equation Modelling (PLS-SEM), a conservative estimate of the recommended sample size for a model in which the maximum number of arrows pointing at a construct is 10 was considered for planning purposes. Accordingly, to achieve a minimum R^2 of 0.10 based on a power of 0.80 and significance level of 0.05 the recommended sample size was 189 while at a significance level of 0.01 was 256. Additionally, rules applied when relying on PLS-SEM that have been considered in IS research is 10 times the largest number of formative indicators used to measure one construct or 10 times the largest number of structural paths pointed at a latent construct in the

structural model (Mikalef & Pateli, 2017). In this regard, the sample size in this research exceeded both requirements.

Data Collection Methods and Instruments

A survey has been a common method to study the phenomenon of the relationship between analytics capabilities and firm performance in other research settings. A survey can also be accompanied with qualitative research such as case studies to make the findings more robust. For example, Mikalef et al. (2016) argued,

> Future studies should empirically test and evaluate this framework by using surveys, interviews, observation, focus groups with experts (e.g., managers, decision makers) and with customers', as well as case studies from the industry. Also, both qualitative and quantitative methods of data collection should be employed. For each different type of data, more than one ways of analysis should be used (e.g., structural equation modelling, qualitative comparative analysis) (p. 7).

I mainly approached data collection using a commonly accepted quantitative method, a theory-based survey, and then additionally sought to develop further understanding with qualitative case studies. Mikalef et al. (2019) argued "A survey-based approach is deemed as an appropriate method of accurately capturing the maturity of firm's big data analytics capabilities" (p. 264). Thus, to capture the main constructs surrounding analytics capabilities and test the relationship with perceived business value, a survey approach is fitting.

Survey research provides quantitative or numeric description of trends, attitudes, or opinions of a population by studying a sample of the population (Creswell & Creswell, 2018). Alternatively, surveys allow measurement of variables by asking questions and examination of the relationship among measures (Singleton & Straits, 2018). Cross-sectional surveys represent a

design in which data is collected on a sample of respondents chosen to represent the target population at one point in time (Singleton & Straits, 2018). This design differs from longitudinal designs in which respondents are questioned over a time period or asked questions at two or more points in time (Singleton & Straits, 2018). Like Gupta and George (2016), I employed a cross-sectional survey to study DAC formation through firm resources and capture business value outcomes framing the period as capturing business experiences in the context of the last three years.

Survey Instruments and Measures

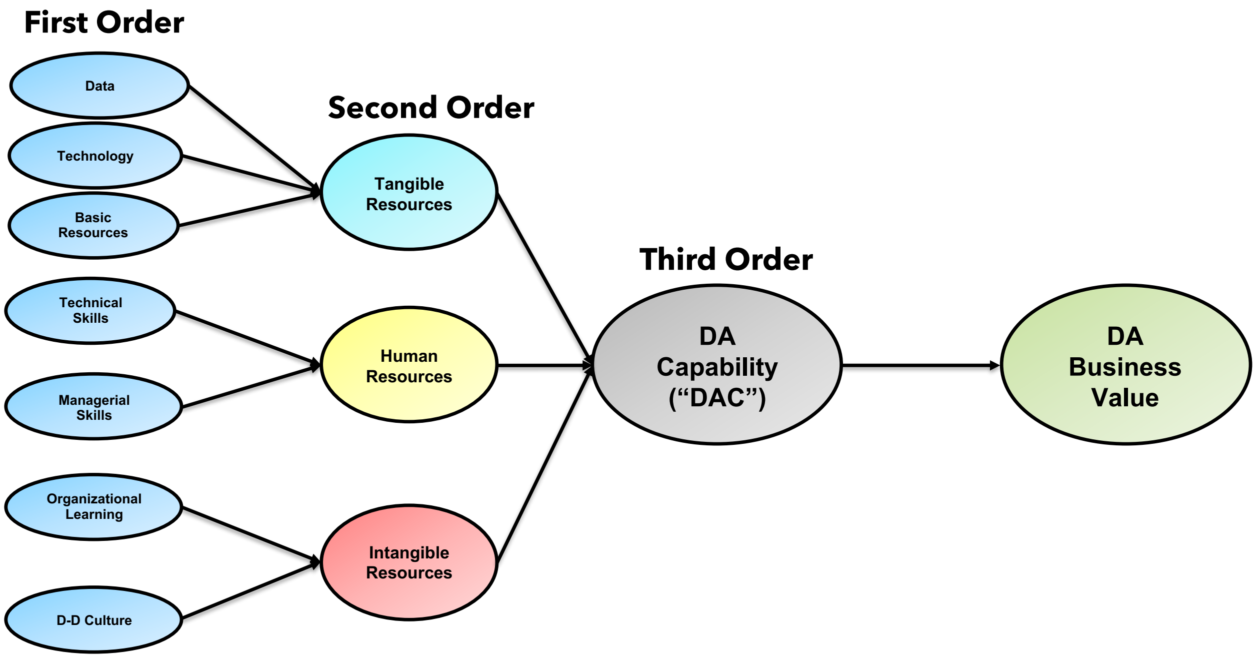
The following provides an overview of the main elements of how the survey items were identified to capture the constructs discussed.

Data Analytics Capability (DAC)

DAC was operationalized as a third-order construct, based on a validated study by Gupta and George (2016). The first-order constructs consisted of more granular items for firm resources data, technology, basic resources, technical skills, managerial skills, data-driven culture, and intensity of organizational learning. The second-order constructs provided a level that ties to the resource-based view of the firm in which there are tangible (data, technology, and basic resources), human (technical skills and managerial skills), and intangible resources (data-driven culture and intensity of organizational learning). These then led to DAC, the third-order construct. While additional indicators were tested, there was a total of 30 DAC-related indicators in the final survey. The questions requested respondents to indicate a value based on agreement level. This corresponded to a 7-point Likert scale (1 = Strongly Disagree to 7 = Strongly Agree). Figure 5 summarizes the hierarchal model including the constructs for formation of a DAC and demonstrates the directional relationship between DAC and business value that was tested.

Figure 5

Hierarchal Model for DAC

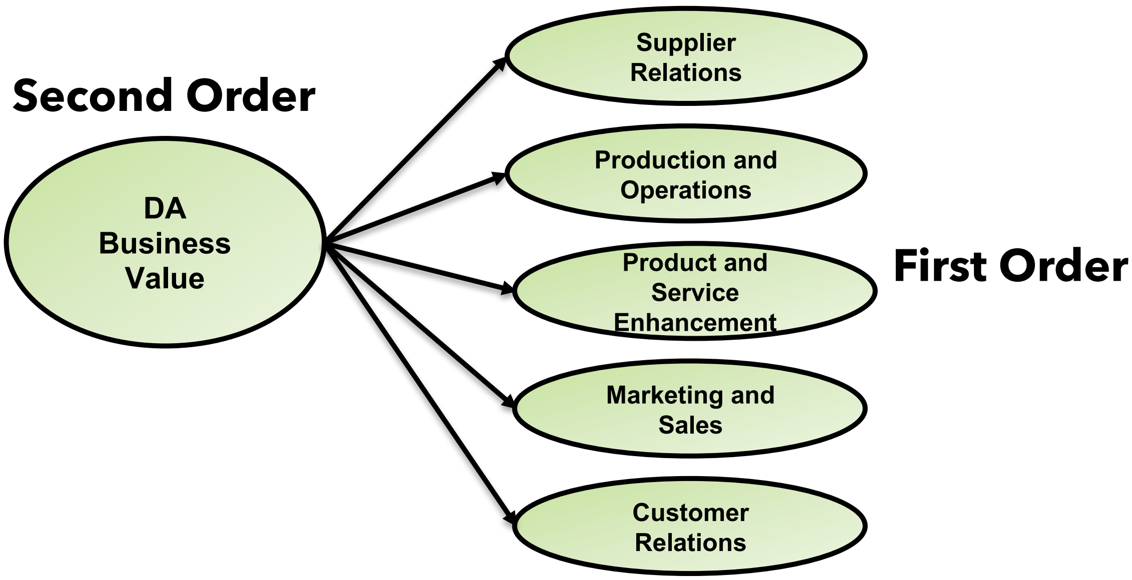


Business Value (BV)

BV is a second-order construct derived from Tallon (2013). The first-order consisted of the different process areas including supplier relations, production and operations, product and service enhancement, marketing and sales, and customer relations. The second-order construct, BV, was produced by these five process areas. This has been validated in several research settings as a valid way of measuring the impact of IT on perceived business performance. There was a total of 24 indicators that captured BV in the final survey. The questions requested respondents to identify a score corresponding to perceived value realized. This corresponded to a 7-point score-based scale (1 = Low Value Realized and 7 = High Value Realized). Figure 6 summarizes how business value is represented as a hierarchal construct, reflected by the five aforementioned process areas.

Figure 6

Hierarchal Model for DABV

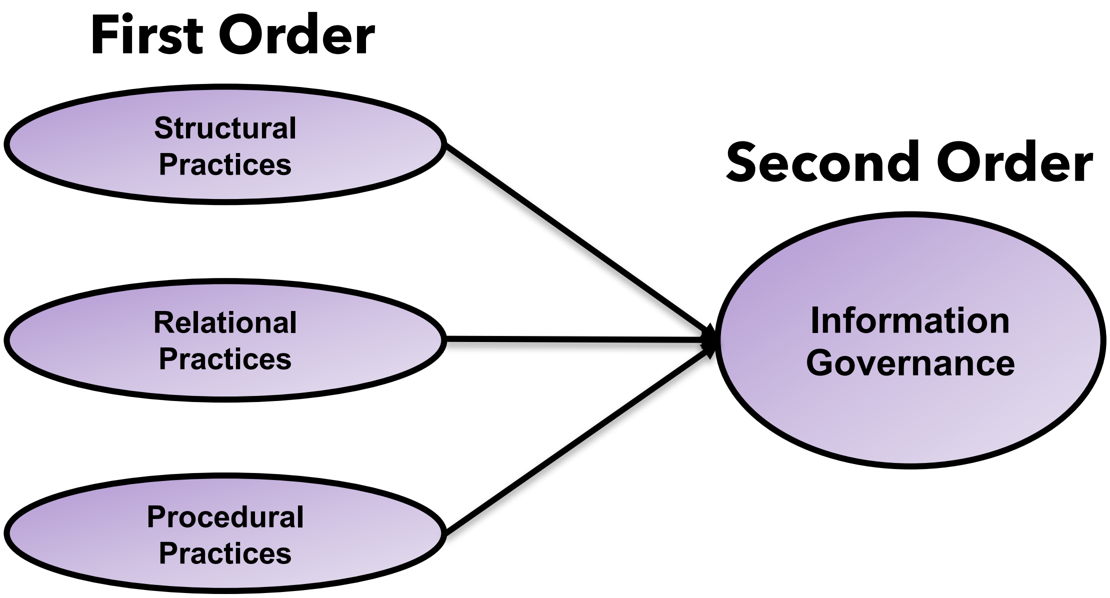


Information Governance (IG)

IG was operationalized as a second-order construct. The first order construct consisted of structural, procedural, and relational practices developed by Mikalef et al. (2019) based on Tallon et al. (2013). These three practices then formed IG. As noted, there has been less attention focused on the role of IG between DAC and performance variables. There was a total of nine indicators that captured IG in the final survey. The questions requested respondents to identify a value based on agreement level. This corresponded to a 7-point Likert scale (1 = Strongly Disagree to 7 = Strongly Agree). Figure 7 summarizes how IG is represented as a hierarchal construct, formed by structural practices, relational practices, and procedural practices.

Figure 7

Hierarchal Model for IG



Instrument Modification

Even though the model and corresponding measurement items had been utilized in previously published empirical research, a pre-test was conducted to test the reliability and validity of the items and the respective constructs or latent variables. According to Ruel et al. (2016), a general rule of thumb is to pilot test a survey on 30 to 100 participants. This process is similar to that reported by Mikalef et. al. (2020c) when utilizing the DAC research model. In addition to testing the statistical properties of the items and constructs, respondents were provided an open response box to provide feedback on the survey. Additionally, they were asked if they would agree to a follow-up interview to clarify any open response feedback.

The purpose of the open response and follow-up interviews was to gain additional insights on SMB respondent comprehension of the questionnaire wording, including data analytics, terms, concepts, and applications. Based on insights from the literature, I suspected

that the survey language would need to be tweaked to capture accurate information around the constructs as understood by SMB owners or key managers. Additionally, gift cards were provided to incentivize participation in a feedback discussion. As a result, I was able to speak to four respondents via phone, going question-by-question to clarify comprehension of the questions and obtain feedback. Ultimately, the information gathered during the follow-up discussions were used to ensure that the questions were clear, scales were interpreted correctly, and the items were capturing what I was attempting to capture. After evaluating the open response feedback, notes from follow-up interviews, and revisiting some of the DAC literature, the instrument was modified for final distribution.

Several items were removed from DAC constructs due to their statistical properties and content. For example, "T1. As a business we have adopted parallel computing approaches (e.g., Hadoop) for data processing and/or analysis" was removed because the concept of database technologies and computing was in fact grouped in the following final survey question: "T4. As a business we have adopted new types of database technologies to process and analyze data such as Hadoop or NoSQL." Another question "TS1. As a business we provide data analytics training to our employees who process or use data" was removed as this question was largely already captured in another question that was part of the pilot and final survey: "TS5. Our data analytics staff are well trained." Gupta and George (2016) excluded a similar question from their analysis based on the statistical properties. Additionally, edits were made to the text of questions to make them easier to interpret and adjustments to the Qualtrics display format were made with the objective of encouraging survey completion.

There were some adjustments which would qualify as noteworthy changes. First, questions regarding environmental factors utilized by Mikalef et al. (2019) were removed due to

many of the constructs demonstrating poor statistical properties, not being part of data analytics resources and therefore not directly contributing to the formation of a capability, and because industry is a more common and efficient way to study the impact of business environment particularly in IS research. Additionally, these items were excluded to make room for important questions related to big data that would allow me to better understand the impact of use of big data versus more traditional data sets.

Ultimately, I felt that if the instrument was too long respondents might suffer from respondent fatigue. Fatigue can cause several issues. For example, a tired or bored respondent may more often answer "I don't know," engage in straight-line responding (i.e., choosing answers down the same column on a page), give more perfunctory answers, or give up answering the questionnaire altogether (Lavrakas, 2008). To mitigate this, survey length was considered.

Based on open response information and interview discussions, more attention was required to address the concept of big data in the context of SMB use and understanding by owners. While the pre-test survey instrument initially provided a pre-questionnaire page with big data concept definitions derived from the literature, I determined that more insights could be gleaned from allowing respondents to self-report the extent that they use big data including the source and frequency of use. Further, this would allow me to tease out differences in business value realized from those that are less or more big data mature or alternatively evaluate maturity of big data use as a control variable when testing the effect of DAC on BV.

As discussed in the literature review, SMBs may be using a lower level of variety of data and software tools. Accordingly, given the objective is to contribute to big data research, to account for different levels of maturity, questions surrounding the source of data were included to capture both the source and frequency of use of such sources. To generate a reliable list, I

consulted George et al. (2014) in which the authors argued “Big data is also a wrapper for different types of granular data” and listed five key sources of high volume data as public data, private data, data exhaust, community data, and self-quantification data (p. 322). While some researchers (Ghasemaghaei, 2021) have relied on the V’s to study big data characteristics and relationships with performance, with the increased availability of low cost, cloud-based storage, the focus for this research was instead on whether the data used is structured versus unstructured, or the variety. For example, Kimble and Milolidakis (2015) differentiated between structured and unstructured data as including both machine generated data and human generated data and argued: “Unstructured data, therefore, is a term that is usually used to describe data whose information content is not readily amenable to automated analysis” (p. 27). Thus, the emphasis in this research was focused on the ability of a SMB business owner and/or staff to readily use and analyze the data source, which I believe is further materialized in capturing the frequency of use.

Finally, a common maturity measure is the level of adoption of software tools or whether the business has graduated from spreadsheets. In both the pre-test and final survey, the frequency of use of data analytics-specific software tools was also captured. Collectively, these two factors can act as controls between a DAC and Business Value and could be further described as “Big Data Maturity Controls”.

Sample Selection and Recruitment

The main source for recruitment was the PCM contact list. A list of 6,721 contacts was provided to me in December 2020. Given that there was no way to tell which of those might fail, I took the existing list and assigned a random number to each contact. It was estimated that approximately 10% of the emails would fail due to inaccurate or closed email addresses. In an initial batch of 300 who were sent an email to participate in the pre-test/pilot survey, 50 (16.7%)

were undeliverable. Since this initial batch was randomly selected, one would expect that only approximately 5,600 of the initial 6,721 contact emails were valid. In any case, an initial recruitment e-mail was sent to the randomly selected sample of 300 contacts. One month later, a recruitment e-mail was sent to a different sample of 200 randomly selected contacts. Finally, a follow-up/reminder e-mail was sent. In this setting, I was able to obtain 28 responses.

Recruitment began in March of 2021. The remaining contacts, who were not contacted for the pre-test, were contacted to complete the final survey. Due to limits on the number of emails that could sent at once, recruitment spanned over five days. Follow-up emails were not sent to build a buffer between the research survey and PCM project recruitment. In mid-April, I began reaching out to personal contacts. In both settings, there were some delays from when contacts received the recruitment message and when they completed the survey. In May 2021, the data was downloaded and no other responses were accepted as useable for analysis. There was a total of 301 fully completed surveys, however 300 were useable (at least 75% of the web-based survey was completed). Assuming the denominator was 5,100 (5,600 minus 500 contacted for the pilot), this suggests that response rate was approximately 6%, not accounting for the fact that a small portion of responses came from recruitment outside of the contact list.

Based on the active study definition for SMBs, the maximum reported employee count was capped to 500. Three respondents who reported 500 were included. This left 300 respondents in the final sample. The recruitment email for the self-administered questionnaire can be found in Appendix B. The items and scales can be found in Appendix C.

Survey Sample Summary Statistics

As a result of the recruitment efforts detailed above, a total of 300 responses were considered. Appendix D contains summary statistics for survey item responses that were analyzed in more detail. Table 2 shows key descriptive statistics of the sample and respondents.

INSERT TABLE 2 HERE

While respondents reported an estimated employee count, firms with less than 10 employees were flagged. This is consistent with the SBA Office of Advocacy standards, "While no definitive definition exists, microbusinesses are defined here as employers with fewer than 10 employees" (Headd, 2017, para 1). Table 2 showed that 43.3% of firms were considered as having less than 10 people, including those that have single operators. Additionally, the reported 2019 revenue ranges showed that the most common estimate for respondent firms was \$100K - \$499K (27% of respondents), \$1M - \$4.99M (22%), and \$1 - \$99K (19%).

There was a wide range of industries represented. The most common were Professional, Scientific, and Technical Services (26.7%), Manufacturing (13.7%), and Other Services (except Public Administration) (9.3%). The survey was targeted toward business owners, as such the most common respondents were a control owner (>50%) who actively operates the business (73.0%) followed by non-control owner (<50%) who actively operates the business (11.7%), and finally, manager or executive with no ownership interest in the business (6.7%). Collectively, almost 85% of the respondents were control or non-control owners who actively operate the business with additional approximately 4% who share control and actively operate the business.

Qualitative Vignettes

Edmunson and McManus (2007) noted that a mixed methods approach in which qualitative data supplements quantitative assessments would enable a more complete explanation of statistical relationships between variables, ensuring the theory provides a valid analysis of a phenomenon. A mixed methods approach "provides a deeper understanding of and rationale for a proposed new construct" (Edmunson & McManus, 2007, p. 1166). This suggests that to address the current research questions, the state of prior theory and research would need to be close to intermediate, rather than advanced, which I agree with. Accordingly, relying on both qualitative and quantitative data can be used to "identify key process variables, introduce new constructs, reconceptualize explanatory frameworks, and identify new relationships among variables" (Edmunson & McManus, 2007, p. 1168). Further, the integration of qualitative methods into IS research has become a necessary step to strengthen quantitative research and findings.

Venkatesh et al. (2013) provide a call for action to employ mixed methods in IS research. Mixed methods can provide a deep understanding of a phenomenon of interest citing the example that "a researcher may use interviews (a qualitative data collection approach) and surveys (a quantitative data collection approach) to collect data about a new IS implementation" (Venkatesh et al., 2013, p. 24). They further cite the diffusion of the internet, the proliferation of numerous network related systems, the availability of devices, and social media as drivers to adopt a mixed method approach where existing research may fall short.

As Creswell and Creswell (2018) noted,

> Another challenge is whether the qualitative samples should be individuals that are in the initial quantitative sample. The answer to this question should be that they are the same individuals because the intent of the design is to follow up the quantitative results and

explore the results in more depth…The idea of explaining the mechanism—how the variables interact—in more depth through the qualitative follow-up is a key strength of the design (p. 222).

It is acceptable to mix case studies with other methods of research. As Yin (2018) noted, “The main investigation may rely on a survey or other quantitative techniques, and your case study may help to investigate the conditions within one of the entries being surveyed” (p. 64). Yin (2018) went on to say, “Interviews are an essential source of case study evidence because most case studies are about human affairs or actions. Well-informed interviewees can provide important insights into such affairs or actions” (p. 119). In addition, with shorter case studies or those that are more focused and expected to take “1 hour or so”, “…a major purpose of such an interview might simply be to corroborate certain findings that you already think have been established…” (Yin, 2018, p, 119). In the context of conducting case studies in addition to quantitative research, Yin (2018) argued,

> …the larger study may have been based on a survey or quantitative analysis or archival data—for example, a study of a households’ financial situations under different income tax conditions. The larger study might then have wanted case studies to illustrate, in greater depth, the experiences of individual families. In this scenario, the questions from your case study might only have surfaced after the survey or archival data had been analyzed, and the selection of the cases might have come from the pool of those surveyed or contained within the archival records (p. 235).

In addition, it is recommended to conduct more than one case study if possible:

> The first word of advice is that, although all designs can lead to successful case studies, when you have the choice (and resources), multiple-case designs may be preferred over

> single-case designs…More important, the analytic benefits from having two (or more) cases may be substantial (Yin, 2018, p. 61).

In this setting, illustrative vignettes, as opposed to full-fledged case studies, were constructed based on follow-up interviews. According to Miles et al. (2014), a vignette is:

> A focused description of a series of events taken to be representative, typical, or emblematic in the case you are studying. It has a narrative, story-like structure that preserves chronological flow and that normally is limited to a brief time span, to one or a few actors, to a bounded space, or to all three (p. 182).

Additionally, with respect to summarizing findings from interviews in a vignette: "…the vignette offers the researcher an opportunity to venture away from scholarly discourse and into evocative prose that remains firmly rooted in the data but is not a slave to it" (Miles et al., p. 183).

Consistent with the above, I sent a recruitment email to nine survey respondents who reported higher levels of business value realized from DA (aside from one respondent), higher levels of data analytics capability resources, and varied in size. Based on this minor recruitment effort, I was able to conduct five interviews. Here, I leveraged a combination of an existing interview guideline utilized by Mikalef et al. (2019) and context from the survey constructs with an emphasis on how these SMBs have been able to generate data analytics-specific business value. The guideline can be found in Appendix E.

CHAPTER 4: DATA ANALYSIS AND FINDINGS

Quantitative Data Analysis Methods

PLS-SEM was used to analyze the data and investigate the relationships in the research model. As a preliminary step, given that the items and corresponding constructs have not been explicitly evaluated in an SMB setting, the reliability and validity of the instrument were assessed by conducting a pre-test run of the survey on a small pilot sample of 28 respondents selected at random from the master PCM database. These assessments combined with qualitative feedback received from interviews conducted with a subset of the respondents to the pilot survey prompted revisions to items in the questionnaire. No modelling was done during the pre-test/pilot phase. Instead, I evaluated the psychometric properties of the survey instrument.

To test the reliability of the survey questions, I calculated and reviewed Cronbach's alpha for each construct. In interpreting Cronbach's alpha, a value of 0.7 to 0.8 is generally acceptable (Field, 2018). Thus, a higher value is better as it indicates that the items within the measures are related. Further, the survey items were tested for validity by examining the correlations and conducting an exploratory factor analysis (EFA).

With respect to validity in survey research, generally the concern is construct validity which addresses how well whatever is purported to be measured has been measured (Lavarkas, 2008). Through factor analysis, one can get a preliminary identification across several survey items if there are groups of items that may represent a smaller set of unobserved, latent variables. More broadly, factor analysis simplifies complex sets of quantitative data by analyzing the correlations between variables to reveal the number of factors which explain the correlations (Jupp, 2006). Additionally, to test the sampling adequacy as a first step prior to factor analysis, a Kaiser-Meyer-Olkin Measure of Sampling Adequacy (KMO) value can be generated and

Bartlett's test of sphericity performed. In a research setting, should the results meet the minimum requirements of a KMO value of larger than 0.5 and close to 1.0 and a statistically significant result for Bartlett's test, then it would be acceptable to conduct a factor analysis as the next step (Field, 2018). In this setting, I reviewed the KMO value, conducted Bartlett's test, and then used PCA with varimax rotation to review the statistical properties of the items and constructs in both the pilot and full study phases. However, the traditional evaluation criteria for both reliability and validity were only emphasized during the pilot phase as PLS-SEM relies on more specific criteria to evaluate the measurement model.

In an initial study performed to build their data analytics capability research model constructs, Gupta and George (2016) noted, "Given that we had data collected on 34 indicators, we first conducted an exploratory factor analysis using principal component analysis and varimax rotation. Seven factors emerged (eigenvalues >1) from this analysis" (p. 7). However, in this dissertation, I was not seeking to develop a new model and identify new factors which then translate to constructs. The intent of this study is to apply a validated research model and corresponding measurement items to a new, understudied setting. However, a major issue with relying heavily on PCA with varimax rotation to dictate latent variables or factors that have already been established in empirical studies is that varimax rotation forces relationships throughout the loading patterns/structures in factor analysis results (Allen, 2017). In this way, the factors associated with theory become increasingly less valid. Thus, while PCA generated general understanding of how the items load to factors should there be no assumed factor structure, there are accepted standards to check the reliability and validity of the measurement model when applying PLS-SEM. For example, in developing the famed Unified Theory of Acceptance and Use of Technology (UTAUT) technology acceptance model, Partial Least

Squares (PLS) was applied to test and examine the reliability and validity of the measures (Venkatesh et al., 2003). Thus, the PCA results in this study were largely used to explore the data. With the constructs already established from previous research and underpinned by theory, it was appropriate to proceed to SEM and evaluate the measurement model using the specified reliability and validity criteria suggested by PLS-SEM scholars.

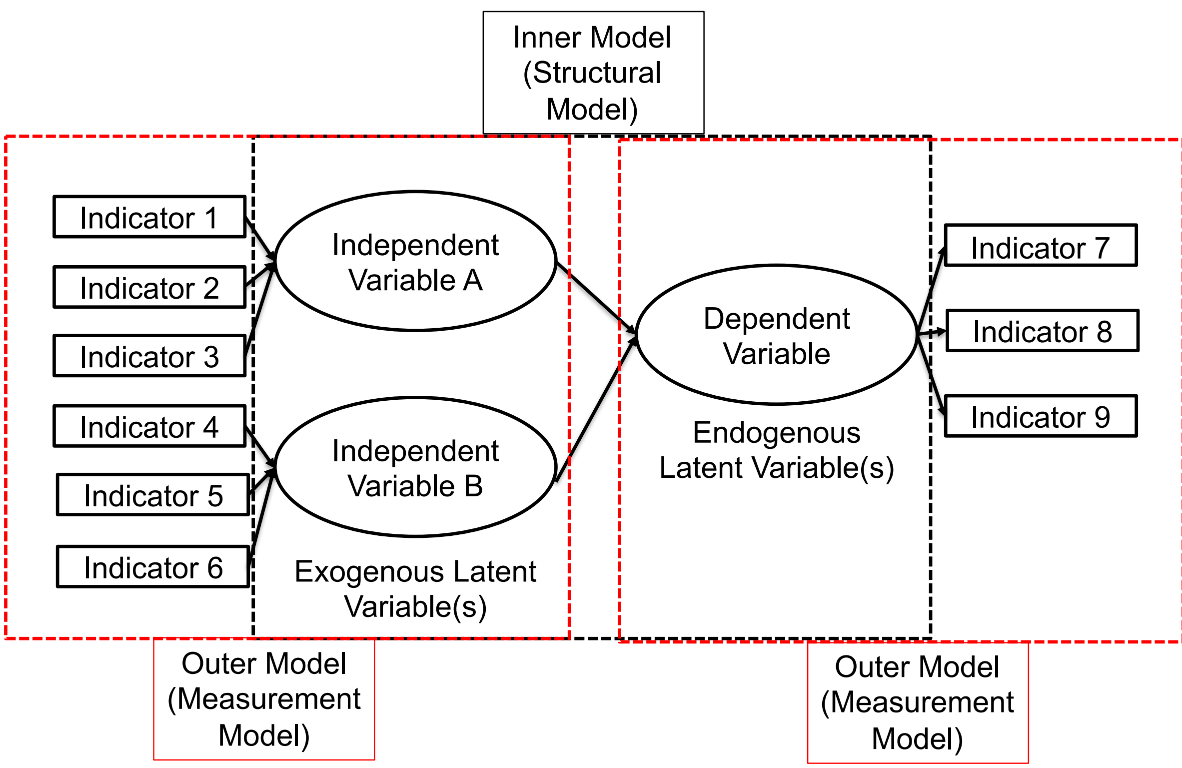
As stated, path analysis in the form of SEM was used to test the structural model and examine cause-and-effect relationships (Singleton & Straits, 2018). More specifically, "PLS-SEM estimates the parameters of a set of equations in a structural equation model by combining an approach similar to principal components analysis with regression-based path analysis" (Hult et al., 2018, p. 5). PLS-SEM is described as a second-generation modelling technique and has been relied upon and developed in the last 20 years to address shortcomings in previously used techniques (Hair et al., 2017). Additionally, PLS-SEM, described as a soft modelling approach, has advantageous properties as it relies on ordinary least squares (OLS) multiple regression, yet no distributional assumptions are made in the construction of the model parameters (Wong, 2019). More specifically, PLS-SEM is a variance-based form of multivariate analysis that allows for the simultaneous analysis of multiple variables that often represent measurements (e.g., measurements associated with companies) obtained from surveys or other primary data (Hair et al., 2017). PLS-SEM is the favored technique if the researcher's goal is theory development and explanation of variance or prediction of constructs (Hair, et al., 2017).

In terms of appearance, a path model is typically shown in diagram form that displays the postulated relationships where constructs (e.g., data-driven culture) are measured by indicators, also called items or manifest variables (e.g., "As a business we consider data a tangible asset" and "As a business we base our decisions on data and analytics rather than on "gut" or instinct")

(Hair et al., 2017, p. 11). The measurement model, also referred to as the outer model, represents the displayed relationship between the items and constructs (Hair et al., 2017). The structural model, also referred to as the inner model, displays the relationship between the constructs, more specifically, the exogenous variables, analogous to IVs in Linear Regression, and endogenous variables, analogous to DVs (Hair et al., 2017). Hair et al. (2018) noted, "Specifically, the PLS-SEM method uses partial regressions to estimate the path coefficients between the latent variables and their indicators in the measurement models, as well as between the latent variables in the structural model" (p. 5). Figure 8, adapted from Wong (2019, p. 10), demonstrates a generic PLS-SEM model and relationships in which the indicators would correspond to questions captured from a survey questionnaire.

Figure 8

Inner vs. Outer Model in a PLS-SEM Diagram



In this research setting, there was only one main IV, DAC, and one DV, business value. However, both DAC and business value are measured through several constructs which themselves have survey questions or items that sought to measure each construct, respectively.

A benefit of PLS-SEM is that it uses proxies to represent constructs that correspond to weighted composites of indicator variables for a given latent variable thereby allowing the process to account for measurement error (Hair et al., 2017). As a result, the weights provided by PLS-SEM allow for additional insights as demonstrated by the following example: "When measuring for example, customer satisfaction, the researcher learns which aspects covered by the individual items are of particular importance for shaping satisfaction" (Hair et al., 2017, p. 16). In fact, using the proxies, "PLS-SEM applies ordinary least squares (OLS) regression with the objective of minimizing the error terms (i.e., residual variance) of the endogenous constructs. In short, PLS-SEM estimates coefficients (i.e., path model relationships) that maximize R^2 values of the (target) endogenous constructs" (Hair et al., 2017, p. 17). Thus, PLS-SEM has the objective of maximizing explained variance in dependent variables in the model and is a variance-based approach (Hair et al., 2017). While the covariance-based approach (CB-SEM) that is commonly used for theory confirmation is seen as an alternative, as outlined by Hair et al. (2017), PLS-SEM should be chosen over such an approach using the following rules of thumb:

- The goal is predicting key target constructs or identifying key driver constructs.
- Formatively measured constructs are part of the structural model.
- The structural model is complex (many constructs and many indicators).
- The sample size is small and/or the data are nonnormally distributed.
- The plan is to use latent variable scores in subsequent analyses. (p. 23).

In this dissertation, the following objectives and characteristics apply to the DAC research model:

- To understand if key resources constructs form an analytics capability and whether this key exogenous construct has a relationship with business value, an endogenous construct.
- Constructs that are both formatively and reflectively measured.
- As a result of their being several orders of constructs (including seven in the first order) associated with the formation of a DAC and several other variables, the model could be characterized as more complex.
- As it is a hierarchal model, it is required that latent scores from lower-order constructs (LOCs) be generated for further use in the analysis in stages.

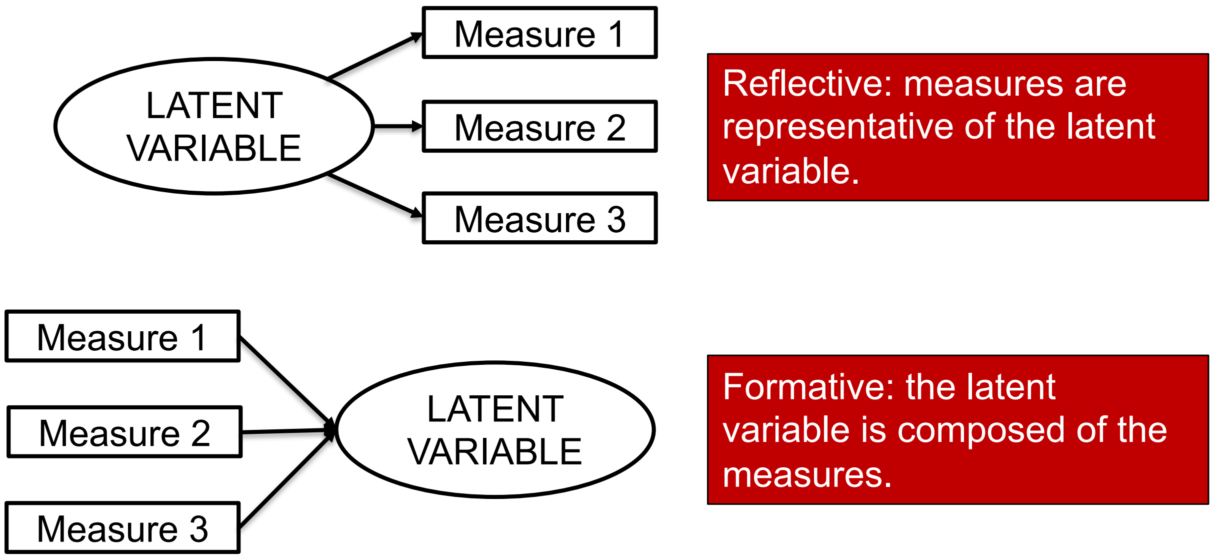
Ultimately, PLS-SEM allowed me to test the relationship between DAC, a higher-order construct (HOC), and DABV, also a HOC. The core research model hypothesizes that DAC has a positive effect on DABV. Finally, Mikalef et al. (2020b) noted on the justification to use PLS-SEM: "Finally, as the proposed research model builds more on exploratory theory building, rather than theory testing, PLS-SEM is a better alternative than covariance-based SEM" (p. 9). While the DAC research model has been applied in different research settings, the model has not been explicitly tested on SMBs or in situations where the respondents, on the majority, have a significant ownership stake. Additionally, in contrast to previous research studies which used more general concepts of performance, perceived business value was measured as a higher-order construct through five value areas. Thus, the purpose of the analysis is to further develop exploratory research findings which further justifies PLS-SEM as an appropriate technique. Now that it has been established that PLS-SEM is suitable for hypothesis testing, several additional

details specific to the complexity and constructs of the DAC research model need to be discussed.

First, as mentioned above, the DAC model includes both formative and reflective constructs. According to Wong (2019), “If the indicators cause the latent variable and are not interchangeable among themselves, they are formative. In general, these formative indicators can have positive, negative, or even no correlation among each other” (p. 11). As a result, I did not need to report indicator reliability, internal consistency reliability, and discriminant validity for formative constructs. Additionally, Garson (2016) argued, “Formative models assume the indicators are "reality" and are all the dimensions of the factor. Dropping an indicator in a formative model is equivalent to dropping a dimension of meaning, causing the meaning of the latent variable to change” (p. 19). In contrast, for reflective constructs, “[if] indicators are highly correlated and interchangeable, they are reflective and their reliability and validity should be thoroughly examined” (Wong, 2019, p. 12). Further, Garson (2016) noted, “Reflective models assume the factor is the reality and measured variables are a sample of all possible indicators of that reality. This implies that dropping one indicator may not matter much since the other indicators are representative also” (pp. 18-19). Figure 9, adapted by Garson (2016, p. 18), shows the directional or line-arrow relationship between the measures or survey items and latent variable or construct for both reflective and formative types.

Figure 9

Illustration of Reflective vs. Formative Constructs



For context, Garson (2016) cited a concrete example for measurement of a hypothetically researched latent variable to show how a construct might be captured reflectively and formatively:

> A reflective model might have the representative measures 'I feel well in this hotel', 'This hotel belongs to my favorites', 'I recommend this hotel to others', and 'I am always happy to stay overnight in this hotel.' A formative model, in contrast, might have the constituent measures, 'The room is well equipped', 'I can find silence here', 'The fitness area is good', 'The personnel are friendly', and 'The service is good' (p. 19).

From this example, one can see a distinction based on the statements of how reflective items can be representative of satisfaction, the formative statements compose a satisfaction construct, and the items inherently represent different elements (i.e., room characteristics, level of noise, fitness

area, staff, service). There are different ways to evaluate the statistical properties of formative and reflective constructs as acknowledged by previous DAC research (Mikale et al., 2020b).

Second, the DAC research model is a hierarchal model that contains both reflective and formative constructs, with the higher orders or all non-first order constructs representing formative relationships. Conceptually, a first-order construct or LOC is the most granular level of characterizing the business requirements for a DAC which in the postulated model include data, technology, basic resources, technical skills, managerial skills, data-driven culture, and organizational learning. Thus, these first-order constructs lead to broader but encompassing constructs that differentiate between different resources consistent with RBV and broader IS literature, including tangible, human, and intangible resources. In this research study, following Gupta & George (2016), all constructs in the first-order were conceptualized as reflective aside from those that comprise tangible resources including data, technology, and basic resources. These were conceptualized as formative as the items represent different facets of tangible resources. For example, basic resources, which has two items, captures adequate funding for DA projects (item BR1) and adequate time provided for DA projects (item BR2). With respect to the DAC research model, Gupta and George (2016) applied four decision rules to identify formative constructs leading to the highest order construct, DAC:

- No single indicator or construct (tangible, human, and intangible resources) can explain the data analytics capability.
- Tangible, human, and intangible resources capture different facets of an organization's data analytics capability.
- Tangible, human, and intangible resources are not required nor expected to co-vary.

- Tangible, human, and intangible resources have different antecedents or constructs that cause them and are distinct business resources.

Third, because of the nature of constructs being different in type, the PLS algorithm computes the composites using different approaches referred to as Mode A for reflective and Mode B for formative. To describe this further, when a reflective measurement model is considered (i.e., relationships from the construct to indicators), the PLS algorithm computes the composites using Mode A in that the outer weights are correlations between the construct and the indicators. In case of a formative measurement model, (i.e., with relationships from the indicators to the construct), the PLS algorithm computes the composites using Mode B or the outer weights are the multiple regression coefficients with the indicators as IVs and the latent variable as the DV. In summary, consistent with previous applications of the DAC model (e.g., Gupta & George, 2016), I modelled technology and basic resource constructs as Mode B, while other first-order constructs were modelled as Mode A. The remaining orders that shape DAC were modeled using Mode B.

As a fourth point, some clarification needs to be made about how the hierarchical model that was developed in this study as it represents a more advanced case of a PLS-SEM model application. In addition to being able to choose between whether a construct is formative or reflective, there are different ways to produce the higher-order construct so that the structural model can be evaluated including the repeated indicators approach or two-stage approaches (Sarstedt et al., 2019). According to Gupta and George (2016), "Following the guidelines of Wetzels et al., the hierarchical model was formally specified, representing the relationships between the indicators, sub-dimensions, and higher-order constructs" (p. 7). Further, according to Wetzels et al. (2009), "As latent variable scores are determinate in PLS path analysis, latent

variables scores for lower-order latent variables can be obtained, which can subsequently be used as manifest variables for the higher-order latent variables" (p. 180). In fact, initial development and subsequent use of the DAC research model with PLS-SEM have relied on the repeated indicator approach. According to Mikalef et al. (2020b):

> A mixture of the repeated indicator approach and a use latent variables scores in a three-stage approach was applied…In the first stage, the repeated indicator approach was used to obtain latent variable scores for the first-order constructs, which in the second stage served as manifest variables in the measurement model of the second-order constructs. This was then repeated for the higher order construct based on latent variables scores of the second-order constructs (p. 10).

While the repeated indicator approach has historically been used to create the higher order latent variables in the context of the DAC model, I argue that it is more appropriate to apply a multi-stage approach.

PLS-SEM experts have found it prudent to use a two-stage approach in settings where a LOC (also called lower-order components) has a formative relationship with a HOC (also called higher-order components) due to the statistical issues that can arise. Matthews et al. (2018) describe, "In this situation, when the relationship from the LOC to the HOC is formative, almost all of the HOC variance is explained by the LOC (R^2 close to 1.0). This can be an issue if there are other relationships pointing to the HOC, as they will have a very small and insignificant impact" (p. 8). The two-stage approach works in the following manner (Riel, et al., 2017):

> The aim of the first stage is to obtain latent variable scores for the first-order constructs. In this first stage of the analysis, the second-order construct is not yet included. It is only in the second stage that the model containing the second-order

> construct is estimated. In the second stage, the scores of the first-order constructs serve as manifest variables of the second-order construct. In essence, the measurement of the first-order constructs is reduced to single items. This reduction is not only useful for statistical reasons (e.g., to avoid multicollinearity among the indicators), but also for practical reasons (e.g., to prevent "double-counting", see Arnett et al., 2003). Most importantly, the two-stage approach allows to place the second-order construct in an endogenous position within the structural model (p. 483).

While there are some options for the several stage approaches, the approach employed in this research more closely resembles the disjoint two-stage approach using the recommended path weighted scheme described by Sarstedt et al. (2019). Accordingly, I used the approach described by Sarstedt et al. (2019):

> Rather than using the repeated indicators approach in stage one, the disjoint two-stage approach considers only the lower-order components of the higher-order construct (i.e., without the higher-order component) in the path model. These are directly linked to all other constructs to which the higher-order construct is theoretically related to. To execute the disjoint two-stage approach, researchers then need to save the construct scores, but only those of the lower-order components…In stage two, these scores are then used to measure the higher-order construct (p. 3).

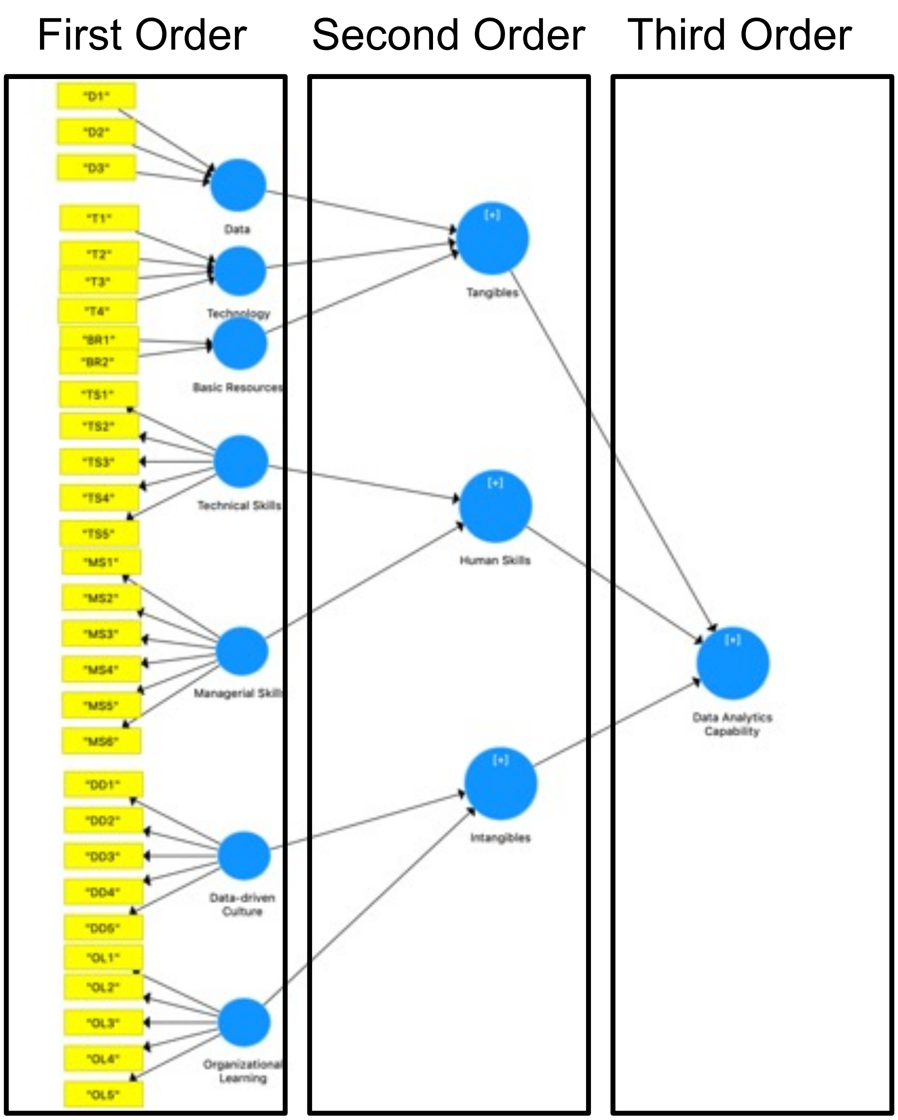
Additionally, the mode should be consistent with the selected approach, as PLS-SEM scholars argued based on a simulation study that "…the disjoint two-stage approach should be estimated

using the standard settings on both stages; that is, Mode A for reflectively specified measurement models and Mode B for formatively specified measurement models” (Sarstedt, et al., p. 4).

Building upon how measured firm resources may form a DAC from a research perspective (i.e., going from survey items to constructs, to a testable, higher-order developed model), a review of the model with annotations that summarize the construct relationships and ultimately the formation of the DAC construct can be seen in Figure 10.

Figure 10

The Data Analytics Capability Model Specified as a Hierarchical Model

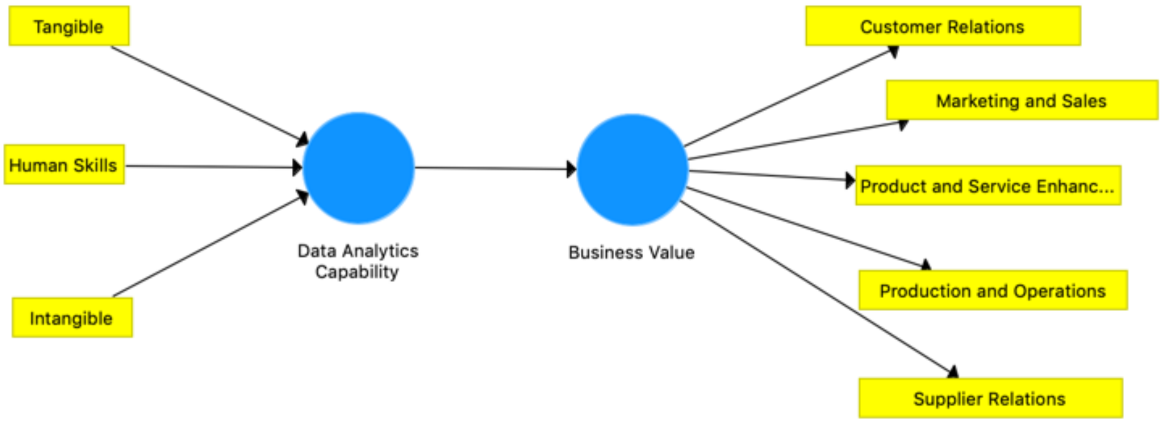


For the first order, the latent scores created from the survey responses to items for each respondent (shown in yellow in Figure 10) that correspond to tangible resources (e.g., data, technology, basic resources) were saved and then used as three items to develop the second order construct, tangible resources. The latent variable scores for the second order items were saved and used as items to form DAC. More specifically, the tangible resources scores were used as an item along with two other items that correspond to the saved scores for human resources and intangible resources, developed at the same stage and manner as the intangible resources construct using the appropriate mode.

The theorized DV, DABV, was also conceptualized as a second-order HOC reflected by the five potential data analytics value areas. In a similar fashion, DABV was specified using a several stage approach with the first order consisting of the five functional/process areas and corresponding measures in which firms may realize business value from DA: customer relations, marketing and sales, product and service enhancement, production and operations, and supplier relations. These were specified as reflective-reflective (i.e., the items reflect each value area and the value area constructs reflect DABV). The basis for this is that I expected firm DA practices to lead to the realization of business value outcomes that would be reflected in the five value areas. The final form of the core research model, at the highest orders, is shown in Figure 11.

Figure 11

The DAC Research Model



Quantitative Data Analysis and Results

Measurement Model: Overview of Evaluation Criteria

The software package SmartPLS3 was used to analyze the survey data. As outlined, PLS-SEM is an appropriate method to analyze the relationships within the postulated DAC research model. Given the hierarchal nature of the model, the starting point is the items and first-order constructs. For simplicity, the criteria were summarized with all orders shown in a single summary where possible below. To assess the measurement model for reflective indicators and constructs, it is advised to assess the following criteria (based on standard measurement model assessment criteria) presented by Sarstedt et al., 2019):

- Internal consistency (Cronbach's alpha, ρA or rhoA, composite reliability)
    - 0.70 to 0.90; 0.60 to 0.70 acceptable in exploratory research (Hair et al., 2017, p. 112); DAC research model: "At the construct level, we examined Composite Reliability (CR) and Cronbach Alpha (CA) values, and established that their values were above the threshold of 0.70" (Mikalef et al., 2020b, p. 9).

- Convergent validity (indicator reliability, average variance extracted)
  - Outer loadings: > 0.70; if >= 0.40 to <= 0.70 then examine the impact of deletion on internal consistency reliability yet can still be retained based on content validity if >= .40 (Hair et al., 2017).
  - Average Variance Explained (AVE): > 0.50 (Hair et al., 2017).
- Discriminant validity
  - Fornell-Larcker criterion: AVE square root of a construct > than highest correlation with any other construct (Hair et al., 2017).
  - Heterotrait-Monotrait ratio (HTMT): While debatable, <= 0.9 has been suggested while 0.85 would be a more conservative value (Hair et al., 2017); < 0.85 (Gupta & George, 2016; Mikalef et al., 2020b).
  - Cross-loadings: an indicators outer loading should be greater than its cross-loading on any other construct (Hair et al., 2017).

Alternatively, for formative indicators and constructs (data, technology, and basic resources that comprise tangible resources in this study) and for the HOCs, tangible, human, and intangible resources that form DAC, it is advised to assess the following criteria (Sarstedt et al., 2019):

- Convergent validity
  - Edwards adequacy coefficient (R2a): calculated by taking the sum of the squared correlations between formative items and their respective formative construct and then dividing the sum by the number of indicators for that construct should be > 0.50 (Gupta & George, 2016; Mikalef et al., 2020b).
- Collinearity between indicators

   - VIF < 10 as a conservative value, or VIF < 3.3 for a more restrictive value is desired (Gupta & George, 2016; Mikalef et al., 2020b).
- Significance and relevance of outer weights (determined using Bootstrapping)
   - If outer weight is significant, there is support to retain the indicator; if not significant, retain if outer loading >= 0.5 or statistically significant (Hair et al., 2017).

As a point of clarification, both outer weights and outer loadings were listed in the evaluation criteria. In fact, they have distinct interpretations particularly in the context of whether the construct is reflectively or formatively specified. Outer weights "are the results of a multiple regression of a construct on its set of indicators. Weights are the primary criterion to assess each indicator's relative importance in formative measurement models" (Hair et al., 2017, p. 323). For example, in a business setting where the focus is marketing activities, "Marketers should pay attention to those indicators with high outer weights as they are the important area or aspect of the business that should be focused on" (Wong, 2019, p. 42). In contrast, according to Hair et al. (2017), outer loadings:

> Are the estimated relationships in reflective measurement models (i.e., arrows from the latent variable to its indicators). They determine an item's absolute contribution to its assigned construct. *Loadings are of primary interest in the evaluation of reflective measurement models but are also interpreted when formative measures are involved* [emphasis added]" (p. 323).

Additionally, "….researchers should also consider a formative indicator's absolute contribution to (or absolute importance for) its construct—that is, the information an indicator provides without considering any other indicators" (Hair et al., 2017, p. 147). Conveniently, standard SmartPLS output produces the PLS-SEM model in diagram form with weights provided for

formative constructs and loadings provided for reflective constructs. Yet, both can be important for interpretation and assessment of criteria. Now that the assessment criteria for the measurement model had been established, the PLS-SEM assessment results from the research sample collected could be discussed.

Each lower order construct was developed in different stages after which the latent scores served as the items for the next order. Thus, the first stage involves using the raw survey items to develop the first-order or lower order construct. SmartPLS conveniently allows the user to save the latent scores using the standard PLS Algorithm. In addition, SmartPLS produces output that allows one to evaluate the measurement model criteria specified above for each stage with some manual calculations required.

Measurement Model: Reflective Constructs

For reflective constructs, it is important to examine the internal consistency, convergent validity, and discriminant validity. In the current study, technical skills and managerial skills that represent human resources, data-driven culture and organizational learning that represent Intangible resources, and, due to business value operationalized as a HOC, customer relations, marketing and sales, product and service enhancement, production and operations, and supplier relations. Additionally, IG was also considered in a secondary analysis as a potential moderator between DAC and DABV. From a measurement relationship perspective, Mikalef et al. (2020c) argued,

> Information Governance (IG) is defined according to the study of Tallon et al. (2013) as a collection of capabilities or practices for the creation, capture, valuation, storage, usage, control, access, archival, and the deletion of information over its life cycle…IG is conceptualized and developed as a second-order formative construct. The three

> underlying pillars that comprise an IG are formulated as first-order reflective constructs (p. 7).

Thus, in this study, IG was considered reflective in the first-order and formative for the second-order, using the first-order scores as items. With respect to reflective constructs, Table 3 summarizes the evaluation of the internal consistency and convergent validity criteria.

Table 3

Reflective - Internal Consistency and Convergent Validity

Construct	Cronbach's Alpha	rho_A	Composite Reliability	Average Variance Extracted (AVE)
Human Skills				
Managerial Skills	0.97	0.97	0.97	0.85
Technical Skills	0.94	0.96	0.96	0.82
Intangible Resources				
Data-driven Culture	0.86	0.87	0.90	0.64
Organizational Learning	0.94	0.94	0.95	0.80
Business Value				
Customer Relations	0.96	0.96	0.97	0.85
Marketing and Sales	0.93	0.94	0.95	0.79
Product and Service Enhancement	0.94	0.94	0.95	0.80
Production and Operations	0.94	0.94	0.95	0.81
Supplier Relations	0.93	0.94	0.95	0.84
Business Value (LOCs as Items)	0.95	0.95	0.95	0.78
Information Governance				
Procedural Practices	0.89	0.89	0.92	0.69
Relational Practices	0.90	0.90	0.95	0.91
Structural Practices	0.64	0.66	0.85	0.74

Outer Loadings

Data Analytics Capability

Item	Managerial Skills	Technical Skills	Data-driven Culture	Organizational Learning
MS1	0.90			
MS2	0.94			
MS3	0.95			
MS4	0.92			
MS5	0.92			
MS6	0.90			
DD1			0.77	
DD2			0.84	
DD3			0.70	
DD4			0.85	
DD5			0.82	
OL1				0.87
OL2				0.91
OL3				0.94
OL4				0.92
OL5				0.83
TS1		0.69		
TS2		0.94		
TS3		0.96		
TS4		0.95		
TS5		0.96		

Business Value

Item	Customer Relations	Marketing and Sales	Product and Service Enhancement	Production and Operations	Supplier Relations
CR_1	0.92				
CR_2	0.91				
CR_3	0.94				
CR_4	0.92				
CR_5	0.92				
MS_1		0.86			
MS_2		0.92			
MS_3		0.87			
MS_4		0.90			
MS_5		0.89			
PO_1				0.92	
PO_2				0.94	
PO_3				0.80	
PO_4				0.91	
PO_5				0.91	
PS_1			0.88		
PS_2			0.89		
PS_3			0.90		
PS_4			0.90		
PS_5			0.91		
SR_1					0.93
SR_2					0.90
SR_3					0.91
SR_4					0.92

Information Governance

Item	Procedural Practices	Relational Practices	Structural Practices
PCR1	0.83		
PCR2	0.76		
PCR3	0.88		
PCR4	0.88		
PCR5	0.79		
RLT1		0.96	
RLT2		0.96	
STR1			0.88
STR2			0.83

Table 3 shows, with respect to the internal consistency, all Cronbach's alpha's were above 0.7 aside from those for structural practices, a first-order construct of IG. However, structural practices was based on only two items. According to Hair et al. (2017):

> Cronbach's alpha is sensitive to the number of items in the scale and generally tends to underestimate the internal consistency reliability. As such, it may be used as a more conservative measure of internal consistency reliability. Due to Cronbach's alpha's limitations, it is technically more appropriate to apply a different measure of internal consistency reliability, which is referred to as composite reliability (p. 111).

Additionally, "When analyzing and assessing the measures' internal consistency reliability, the true reliability usually lies between Cronbach's alpha (representing the lower bound) and the composite reliability (representing the upper bound)" (Hair et al., 2017, p. 112). In evaluating the composite reliability for structural practices, the value of 0.85 meets acceptable internal consistency standards as "values between 0.70 and 0.90 can be regarded as satisfactory" (Hair et al., 2017, p. 111). With respect to convergent validity, all AVE values were above the specified 0.5 cut-off (Hair et al., 2017).

Additionally, in assessing indicator reliability, all but one item loading (TS1: "As a business we hire new employees who already have the data analytics skills and experience we need"), was not above 0.7. However, it nearly met 0.7 with a value of 0.69. As discussed by Hair at al. (2017), one should consider deletion if the value is between 0.4 and 0.7 after analyzing the impact on internal consistency reliability. With TS1 included, the internal consistency reliability was high, with a Cronbach's alpha of 0.94 and composite reliability of 0.96. However, when excluded, Cronbach's alpha increased to 0.97 and composite reliability increased to 0.98. Yet, a composite reliability of over 0.95 may not be desirable as it could negatively impact the content

validity (Hair et al., 2017). Additionally, there is a theoretical justification that inclusion of TS1 which captures hiring of skilled, data analytics staff is important in measuring the resources associated with the technical skills construct. Therefore, TS1 was retained.

Next, an assessment of the discriminant validity for reflective constructs was performed. The result of this analysis is summarized in Table 4, where the square root AVE is shown on the diagonal and construct correlations displayed otherwise.

Table 4

Reflective: Discriminant Validity

Discriminant Validity Fornell-Larker criterion Construct	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
(1) Basic Resources															
(2) Data	0.44														
(3) Data-driven Culture	0.55	0.53	0.80												
(4) Managerial Skills	0.64	0.43	0.62	0.92											
(5) Organizational Learning	0.44	0.24	0.52	0.55	0.89										
(6) Technical Skills	0.64	0.41	0.55	0.83	0.51	0.90									
(7) Technology	0.47	0.63	0.55	0.52	0.32	0.51									
(8) Customer Relations	0.43	0.49	0.55	0.48	0.36	0.39	0.49	0.92							
(9) Marketing and Sales	0.41	0.47	0.57	0.48	0.34	0.39	0.48	0.84	0.89						
(10) Product and Service Enhancement	0.41	0.53	0.56	0.52	0.37	0.44	0.54	0.82	0.74	0.90					
(11) Production and Operations	0.40	0.49	0.57	0.51	0.35	0.40	0.52	0.81	0.77	0.82	0.90				
(12) Supplier Relations	0.45	0.52	0.56	0.48	0.33	0.40	0.49	0.74	0.73	0.73	0.77	0.91			
(13) Procedural Practices	0.50	0.48	0.57	0.55	0.48	0.50	0.49	0.59	0.52	0.57	0.56	0.52	0.83		
(14) Relational Practices	0.38	0.43	0.49	0.55	0.34	0.50	0.47	0.51	0.43	0.52	0.53	0.43	0.63	0.91	
(15) Structural Practices	0.53	0.44	0.60	0.62	0.42	0.58	0.45	0.51	0.50	0.56	0.54	0.45	0.61	0.66	0.86
AVE square root of a construct > **highest correlation with any other construct**			Yes	Yes	Yes	Yes		Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

As shown in Table 4, the AVE square root of each construct was greater than its highest correlation with any other construct, supporting discriminant validity.

Given that separate construct scores were developed in stages for DAC, BV, and IG, to avoid double counting, as would occur if the repeated indicator approach had been employed, Heterotrait-Monotrait ratio (HTMT) values and cross-loadings were produced and evaluated separately for each broader construct. Hair et al. (2017) noted,

> The HTMT approach is an estimate of what the true correlation between two constructs would be, if they were perfectly measured (i.e., if they were perfectly reliable). This true

correlation is also referred to as disattenuated correlation. A disattenuated correlation between two constructs close to 1 indicates a lack of discriminant validity (p. 118).

Appendix F shows all HTMT all values associated with the DAC construct were less than 0.9, while only managerial skills and technical skills would not satisfy the most conservative cut-off of 0.85. Similarly, in evaluating BV, all HTMT values were less than 0.9, while several construct relationships, customer relations, marketing and sales, production and operations, and product and service enhancement, would not satisfy the most conservative cut-off. The same holds for IG as structural practices and relational practices that would marginally not satisfy the most conservative cut-off with a HTMT value of 0.86. Given that 0.85 is a conservative cut-off and all HTMT values met the 0.9 threshold, the HTMT discriminant validity criteria were largely satisfied. Also, with respect to discriminant validity, as shown in Appendix G, all outer loadings for items are greater than their cross loading for any other construct in their respective HOC. Collectively, the psychometric properties for the reflective constructs were evaluated and satisfy accepted criteria, though in some cases more moderate criterion thresholds were applied.

Measurement Model: Formative Constructs

After reflective constructs were assessed, the formative constructs were formally assessed. For formative constructs, it is important to examine convergent validity, collinearity (as correlation is not expected), and the significance and relevance of outer weights. With respect to the formative constructs, Table 5 summarizes the evaluation criteria.

Table 5

Formative: DAC Convergent Validity, Collinearity, and Outer Weights

Convergent Validity, Collinearity between Indicators, Outer Weights, and Outer Loadings Data Analytics Capability and Information Governance						
Construct/Item	**Outer Weights**	**Outer Weights Significance (p-value)**	**VIF**	**R2a (Edwards Adequacy Coefficient)**	**Outer Loadings**	**Outer Loadings Significance (p-value)**
Data				0.79		
D1	0.32	0.002	1.51		0.76	0.000
D2	0.70	0.000	1.66		0.95	0.000
D3	0.14	0.191	1.49		0.67	0.000
Technology				0.76		
T1	0.33	0.000	1.64		0.79	0.000
T2	0.20	0.028	1.59		0.72	0.028
T3	0.33	0.000	1.32		0.71	0.000
T4	0.43	0.000	1.56		0.83	0.000
Basic Resources				0.93		
BR1	0.38	0.003	2.18		0.89	0.000
BR2	0.68	0.000	2.18		0.97	0.000
Tangible Resources				0.77		
Data	0.08	0.327	1.71		0.62	0.000
Technology	0.36	0.000	1.79		0.76	0.000
Basic Resources	0.73	0.000	1.35		0.93	0.000
Human Skills				0.95		
Managerial Skills	0.67	0.000	3.30		0.98	0.000
Technical Skills	0.37	0.000	3.30		0.93	0.000
Intangible Resources				0.87		
Data-Driven Culture	0.67	0.000	1.38		0.92	0.000
Organizational Learning	0.47	0.000	1.38		0.82	0.000
Data Analytics Capability				0.87		
Tangible Resources	0.47	0.000	2.22		0.89	0.000
Human Resources	0.12	0.328	2.44		0.81	0.000
Intangible Resources	0.54	0.000	2.04		0.91	0.000
Information Governance				0.86		
Structural Practices	0.38	0.000	1.99		0.86	0.000
Relational Practices	0.20	0.045	2.09		0.80	0.000
Procedural Practices	0.56	0.000	1.87		0.92	0.000

Table 5 shows that most items and constructs had statistically significant outer weights. The exceptions were the items for D3 (As a business we combine external data with internal data), the lower-order constructs data (first-order component of DAC), human resources (second-order component of DAC), and structural practices (first-order component of IG). According to Wong (2019), one should “only remove the indicator if both of its outer weights and outer loadings are not significant” (p. 42). Additionally, Hair et al. (2017) noted, “When an indicator’s outer weight is non-significant but its outer loading is high (i.e., above 0.50), the indicator should be interpreted as absolutely important but not as relatively important. In this situation, the indicator

should generally be retained" (p. 148). Consistent with previous applications of the DAC research model, the items and constructs were kept on the basis that the indicators and constructs capture different facets of data, tangible resources, DAC, and IG, respectively, that demonstrate absolute importance as shown by the outer loading size and statistical significance. For example, Mikalef et al. (2020b) noted with insignificant weights associated with items in the technology construct in their research model:

> Since the technology construct is proposed as an aggregate of three items, where each captures a different big data-related technology, we believe that it is critical to include the indicator in the model as it makes a distinct contribution. A similar approach is followed by Gupta and George (2016) in their operationalization of BDAC (p. 9).

Indeed, Sarstedt et al. (2019), when presenting an example evaluation of a hierarchal model, kept an insignificant construct (lower-order) in a measurement model on the basis that the construct's loading was larger than 0.5. All the outer loadings for the formative indicators and constructs met the threshold criteria and were statistically significant.

In relation to multi-collinearity, all VIF values were less than 3.3 except for the value for managerial skills and technical skills which were approximately 3.3 (3.297). Given it was much lower than 10, this was not a cause for concern. Finally, with respect to convergent validity, all calculated R2a values were well above 0.5.

Having investigated and confirmed the psychometric properties for the measurement model, the structural model was assessed.

Structural Model

Given that the psychometric properties of the DAC research model and BV have been established at this point, the relationship between DAC and DABV could be evaluated. To assess

the structural model, it is advisable to evaluate collinearity between constructs, significance and relevance of the path coefficients, and explanatory (R2) and predictive power (Sarstedt et al., 2019). Path coefficients guide interpretation as these “are estimated path relationships in the structural model (i.e., between the constructs in the model). They correspond to standardized betas in a regression analysis” (Hair et al., 2017, p. 323). The closer the path coefficient value is to +1, the stronger the positive relationship. Alternatively, path coefficients closer to -1 suggest a strong the negative relationship (Hair et al., 2017). In addition, Hair et al. (2017) describes,

> R2 values: are the amount of explained variance of endogenous latent variables in the structural model. The higher the R2 values, the better the construct is explained by the latent variables in the structural model that point at it via structural model path relationships. High R2 values also indicate that the values of the construct can be well predicted via the PLS path model (p. 326)

With respect to the evaluation criteria, “…the disjoint two-stage approach uses multi items in the second stage, which permits the application of all structural model assessment criteria. Hence, when using the disjoint two-stage approach, researchers should assess the structural model on the grounds of stage two results” (Sarstedt et al., 2019, p. 5).

While the latent scores and measurement model criteria were assessed using the standard PLS algorithm with the path weighting scheme and statistical significance was tested using bootstrapping (n = 5000), the final model was run using PLSc. According to Wong (2019),

> The traditional PLS algorithm has its shortcomings. Dijkstra and Schermelleh-Engel (2014) argue that it overestimates the loadings in absolute value and underestimates multiple and bivariate correlations between latent variables. It also found that the R^2 value of endogenous latent variables is often underestimated (Dijkstra, 2010) (p. 116).

Additionally, Wong (2019) states that,

> Building on Nunnally's (1978) famous correction for attenuation formula, the Consistent PLS (PLSc) is proposed to correct reflective constructs' correlations to make estimation results consistent with a factor-model (Dijkstra, 2010; Dijkstra, 2014; Dijkstra & Henseler, 2015a; Dijkstra & Schermelleh-Engel, 2014) (p. 117).

Wong (2019) noted that if there is a mixture of reflective and formative constructs it is recommended to use Consistent PLS and bootstrapping. More importantly, "There are also other considerations when using PLSc. For example, if a researcher's model utilizes a higher-order construct, he or she should just use the two-stage approach and not the repeated indicator approach as the latter does not work well with PLSc" (Wong, 2019, p. 117). Additionally, the correction only applies to reflectively measured constructs. Thus, PLSc was implemented to produce the base model results which allowed for evaluation of the standard, structural model assessment criteria.

Prior to analyzing the relationship between DAC and BV, an initial model which consisted of control variables based on elements that were collected in the survey instrument was conducted. In other words, prior to evaluating the effect of DAC on BV, all the controls were included as IVs to see how large of an effect DAC would have or how much additional variation in BV could be explained by DAC.

Several variables were created including a dummy variable that compared firms that comprised of less than 10 individuals versus those that were reported to be comprised of 10 or more, a dummy variable for whether the respondent was a control owner versus non-control owner, a dummy variable for industry (e.g., product or movement of goods industry versus other), and a dummy variables that correspond to comparisons for firm-level strategies (e.g.,

operational excellence, customer intimacy, and product and service leadership). In terms of product movement industries, the following were considered based on the NAICS primary industry code:

- 11 Agriculture, Forestry, Fishing and Hunting
- 21 Mining, Quarrying, and Oil and Gas Extraction
- 23 Construction
- 31-33 Manufacturing
- 42 Wholesale Trade
- 44-45 Retail Trade
- 48-49 Transportation and Warehousing

With respect to firm strategy, firms were classified as operationally excellent (OE), customer intimate (CI), product and service leaders (PSL), or mixed foci based on whether the firm provided 50 or more points to a value discipline. Also, to test if there were differences across when the response occurred, the first 100, middle 100, and last 100 respondents were compared. With these variables in the model, the R^2, or proportion of variance in BV explained by the control variables, was only 5.3%. Yet, when DAC was added as an IV to the model, the R^2 increased to 45.9%. Thus, DAC has a standardized path coefficient ($\beta = 0.67$) that is much larger than any control variable and increases the proportion of variance substantially.

With respect to effect size, Wong (2019) noted,

> …there can be a detailed discussion of the model's f2 effect size which shows how much the exogenous latent variable contributes to an endogenous latent variable's R2 value. In simple terms, effect size assesses the magnitude of relationship between the latent

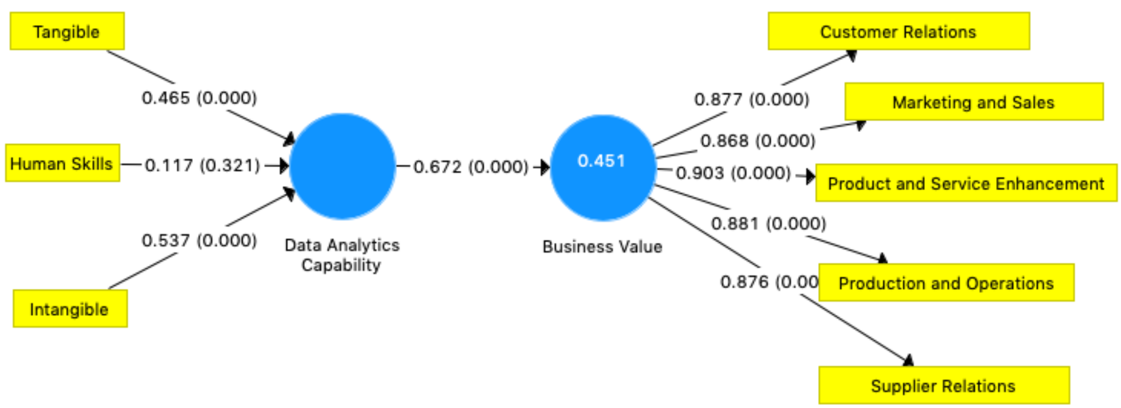
variables. Such discussion can be important because effect size helps the researchers to assess the overall contribution of a research study (p. 38).

According to Wong (2019), "Effect size of 0.02, 0.15, and 0.35 indicates small, medium, and large effect, respectively" (p. 38). The effect size when DAC was included was large, 0.75. In other words, the variance accounted for or attributed by DAC is large. Based on the strength of DAC, it was appropriate to run the model with DAC only to analyze the relationship between DAC alone and BV.

As shown in Figure 12, there is a statistically significant, positive effect of DAC on BV.

Figure 12

Results for the Base DAC and BV Model



The base research model shows that DAC has a positive effect on DABV and accounted for 45.1% of the variance in DABV. The path coefficient which measures the effect ($\beta = 0.67$) and statistical significance ($p < 0.001$) are shown in Figure 12. In other words, a one-unit increase in the standard deviation in DAC corresponds to a 0.67 standard deviation increase in DABV. In this setting, the effect size was 0.82, which suggests a large effect when comparing a model without any variables to the model that includes DAC.

In terms of predictive relevance, it is advised to evaluate Stone-Geisser's (Q2) values "(i.e., cross-validated redundancy measures)" using Blindfolding in which "an omission distance of (OD) of 5 to 10 is suggest most research" (Wong, 2019, p. 39). More specifically, the Q2 value is interpreted in the following manner according to Hair et al. (2017):

> This measure is an indicator of the model's out-of-sample predictive power or predictive relevance. When a PLS path model exhibits predictive relevance, it accurately predicts data not used in the model estimation. In the structural model, Q2 values larger than zero for a specific reflective endogenous latent variable indicate the path model's predictive relevance for a particular dependent construct (p. 207).

In the current study, the Q2 value was 0.347 indicating predictive relevance.

In addition to the path coefficients and main structural criteria, one can evaluate the weights (formative) and loadings (reflective) associated with each of the lower order constructs that appear in the structural model function as items. The weights are shown for formative constructs or resources that form DAC (tangible, human skills, and intangible), while the loadings are shown for reflective constructs for DABV (customer relations, marketing and sales, product and service enhancement, production and operations, and supplier relations) along with p-values corresponding to significance of the weights and loadings, respectively. Review of the weights is useful to understand the managerial implications of the results.

As such, both tangible and intangible resources have a fairly equal relative contribution to DAC as when the relative importance is compared, the ratio of intangible (the construct with the highest relative importance) to tangible was approximately 1.15 (0.54/0.47), while human skills relative contribution was weak and not significant as the ratio of intangible to human skills was approximately 4.59 (0.54/0.12). However, the outer loading was over 0.8 (well above 0.5) and

thus it is a good construct as it relates to the formation of DAC. This is further justified by theory or the notion that human skills are necessary to analyze and make sense of data and is important.

For SMBs, however, tangible, intangible, and human skills were all important (absolute based on the loading) in contributing to the formation of a DAC. Tangible and intangible resources were of much more significance and importance to SMBs (relative based on the weight). Thus, SMB owners who are interested in elevating their DAC should consider developing the intangible resources (i.e., data-driven culture, intensity of organizational learning) and tangible resources (i.e., data, technology, basic resources). In evaluating the outer loadings related to DABV, all five value areas demonstrated high loadings and were statistically significant ($p < 0.001$), suggesting all are strong constructs for BV.

IG as a Moderator

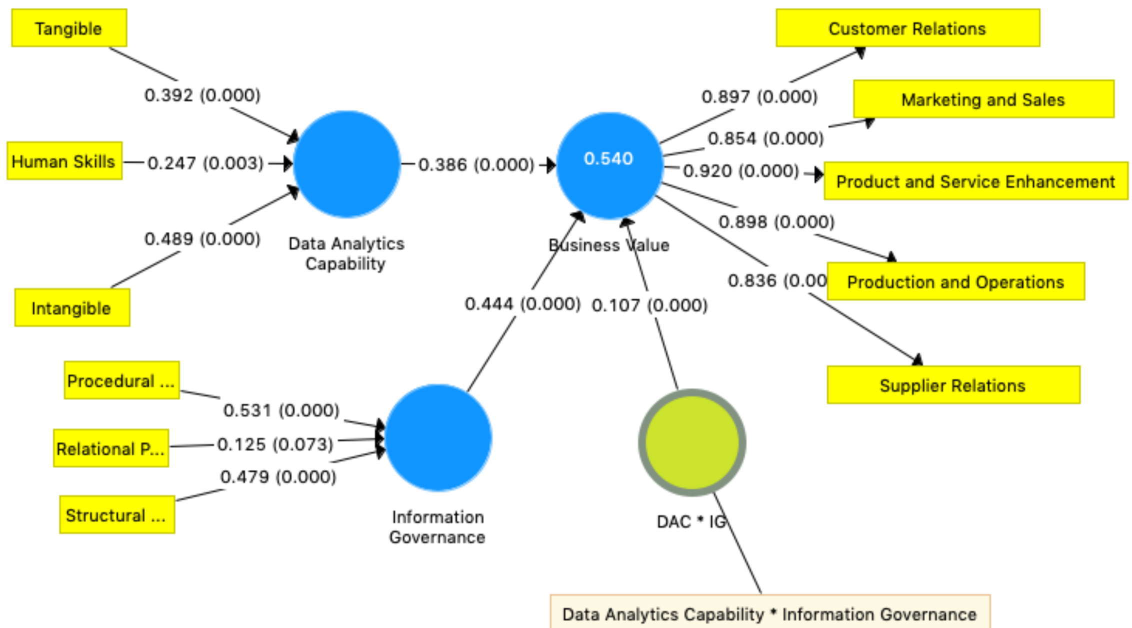
Another factor that has been highlighted to be of potential importance in realizing value from DA is IG. More specifically, IG has been hypothesized to enhance the effect of DAC on DABV realized. To evaluate the potentially moderating effect of IG, the construct was operationalized in a consistent manner of other second-order constructs presented in which the items associated with structural, procedural, and relational practices were considered as reflective of such practices and then the first-order, latent variables scores were saved and used in a formative fashion as items for the second-order IG construct.

The moderator was established using a two-stage approach, as "Chin et al. (2003) proposed the two-stage approach as a means to run a moderation analysis when the exogenous construct and/or the moderator are measured formatively" (Hair et al., 2017, p. 251). Additionally, "When interpreting [the] results of a moderation analysis, the primary interest is with the significance of the interaction term. If the interaction term's effect on the endogenous

construct is significant we conclude that the moderator M has a significant moderating effect on the relationship…" (Hair et al., 2017, p. 256). Figure 13 presents the model results.

Figure 13

Results for DAC and BV Model when Including IG as a Moderator



As shown in Figure 13, when incorporating IG as a moderator, the interaction term shows a statistically significant and positive effect. In other words, the higher the level of IG practices present, the stronger the relationship between DAC and DABV, that is IG strengthens the relationship between DAC and DABV.

The estimated value of the DAC β is described as the simple effect and signifies the strength of the relationship between DAC and BV when IG (the moderator variable) has a value of zero. If the level of IG is increased by one standard deviation, the simple effect of .39 is expected to change by 0.11, the interaction β. Thus, the relationship between DAC and BV would be expected or estimated to change to a value of $0.39 + 0.11 = 0.50$, if (ceteris paribus) the average or mean value of the moderator increases by one standard deviation (Hair et al., 2017).

Further, one can compare the change in R-squared and the corresponding effect size (f2). Comparing the R-squared of 54.0% when the moderator is included to 45.1% (without the moderator), the variance explained increased by almost 9% resulting in an effect size of 0.19, which would be classified as a medium effect.

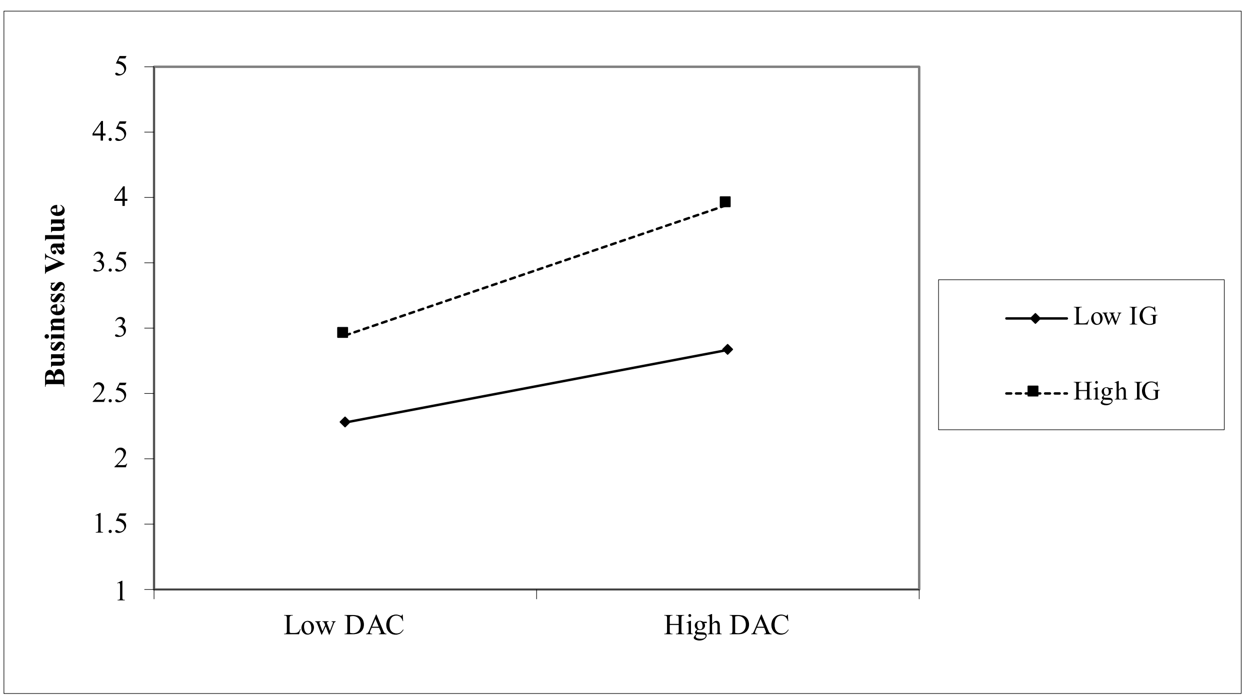
Additionally, as Ramayah et al. (2018) noted

> The size and precise nature of this effect is not easy to define from examination of the coefficients alone, and it becomes even more so when one or more of the coefficients are negative, or when the standard deviations…are very different (Dawson, 2014). Thus, Dawson (2014) suggested that to follow up for the significant interactions, an interaction plot can be drawn (p. 43).

The interaction plot is shown in Figure 14.

Figure 14

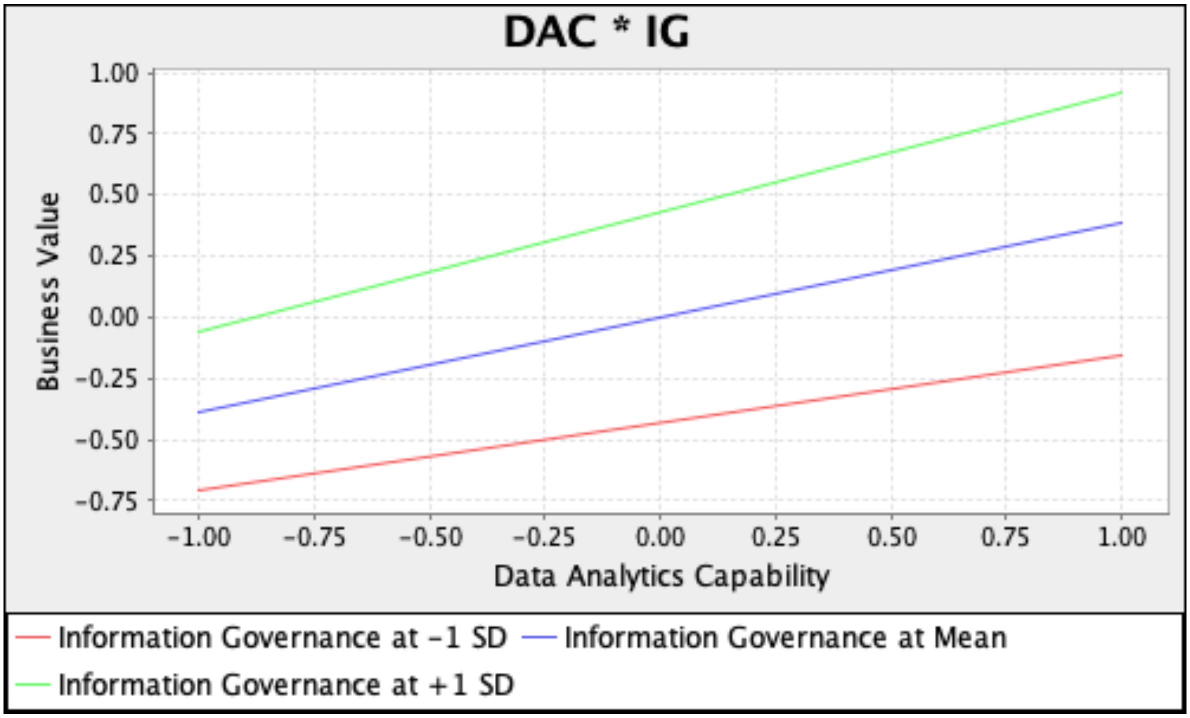
Interaction Plot: IG as a Moderator



Based on the interaction plot, the line ‘High IG’ has a steeper slope when compared to the ‘Low IG’ line, suggesting that the positive relationship between DAC and BV is stronger when IG is high. An alternative is to examine the simple slope analysis which is shown in Figure 15.

Figure 15

Simple Slope Analysis: IG as a Moderator



In Figure 15, the middle line represents an average level of the construct IG. The upper line represents a high level of the moderator construct (mean + 1 standard deviation) while the lower line represents a low level of IG (mean - 1 standard deviation). From the positive slopes shown, DAC has a positive effect on BV. In addition, the high-level line has a steeper slope when compared to the low level line, hence a consistent conclusion that the presence of IG leads to a stronger, positive relationship between DAC and BV.

Revised Model with Big Data and Software Use Frequency

It was proposed that the maturity level of big data could be used as a control to determine if the volume, variety, and veracity of data available has a statistically significant impact on DABV. Additionally, the level of sophistication of software tools being used can also act to influence (control) the relationship between DAC and DABV, and similarly provide an assessment of the maturity of the firm in terms of being able to process and analyze data. Both variables were measured on a Likert scale with respondents being asked to report on the organizations frequency of use of each data and software category. To develop a score for each variable, the sum of values for each category was calculated. For the big data score, only sources of data which were considered unstructured and would not be classified as traditional or easy to analyze were included in the aggregate score. These are summarized below:

- Data streaming from sensors, RFIDs, Internet of Things (IoT), machines and/or vehicles
- Data from social media postings, online customer reviews/ratings on third party websites, and online community conversations
- Data generated from wearable devices (e.g., employee biometric or movement tracking)
- Website browsing and search data (e.g., cookies) from computers and/or mobile devices
- Geographic location data (e.g., customer or employee location) from computers and/or mobile devices
- Cross-app messaging and communication data from computers and/or mobile devices

However, for the software tools score, all items were included. These are shown below:

- Custom dashboard reports manually produced by you or an employee of your business (i.e., produced using Tableau, Microsoft Power BI, Looker, Alteryx, Tableau, Microsoft Power BI, Google Data Studio, or another application)

- Generic dashboard reports that are generated automatically by a Software as a Service (SaaS) provider or mobile application (i.e., Salesforce or another software application your company subscribes to)
- Statistical programming software used by you or an employee of yours to analyze and/or visualize data (i.e., SAS, SQL, SPSS, STATA, R, Python, or another programming language or advanced tool used for statistical analysis)
- Microsoft Excel or another standard spreadsheet software used by you or an employee of yours to analyze and/or visualize data
- Other Data Analytics tools and/or software not identified above

Table 6 presents summary statistics associated with the sum score values for all data, big data only, and software tools:

Table 6

Summary Statistics of Sum Scores used as Controls

Statistic	Big Data Score (Max = 42)	Software Tools Score (Max = 35)*
N	300	300
Min.	6.00	0.00
Max.	42.00	35.00
Median	14.00	15.00
Mean	16.38	15.85
Std. Dev.	8.35	7.37

Note. 3 respondents who did not report on software tools were assigned a score of 0.

As shown in Figure 16, there was a positive and significant effect of both big data use and software tools use.

Figure 16

Results for DAC and BV Model when Controlling for Big Data Use and Data Analytics Software Tools Use

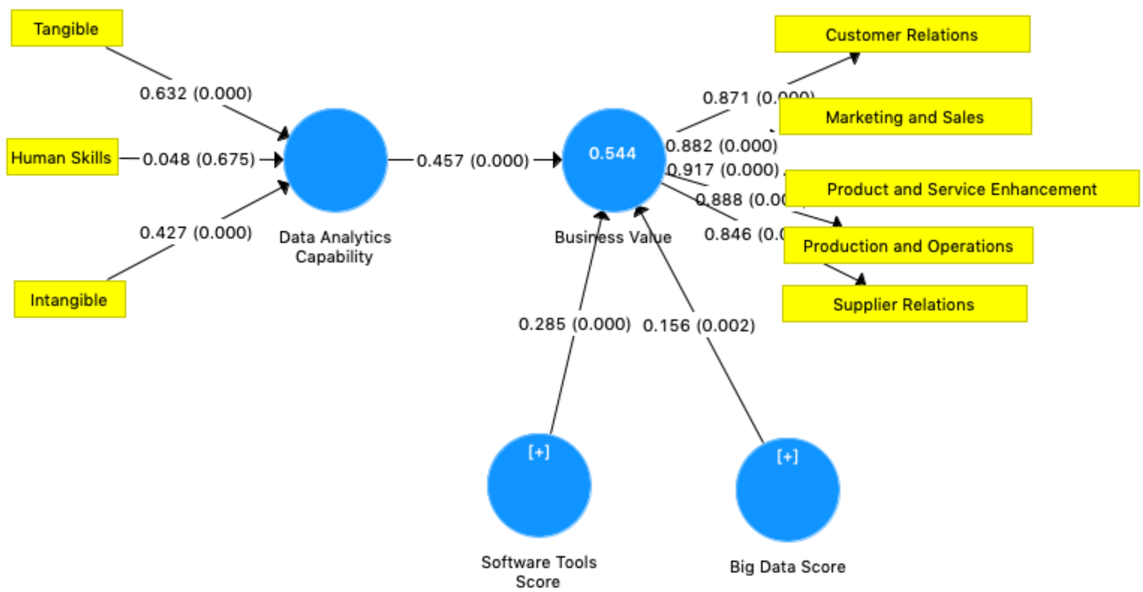


Figure 16 shows that when controlling for big data and software use, or big data maturity, the variables explained 54.4% of the variance in BV. Additionally, all had a positive effect on DABV, with DAC having the largest impact ($\beta = 0.46$, $p < 0.001$), followed by software tools ($\beta = 0.29$, $p < 0.001$), and big data ($\beta = 0.16$, $p < 0.01$). Additionally, VIF values of 1.25, 1.47, and 1.64 for big data score, DAC, and software tools, respectively, indicated that multicollinearity was not an issue. This is an important finding at it suggests that stronger reliance on and adoption of more advanced software tools and big data have a positive impact on DABV, and their inclusion as a complement to DAC helps explain a larger proportion of the variance in DABV.

Multi-group Analysis (MGA)

MGA allows the researcher to compare subgroups or subsamples to determine if there are statistically significant differences across groups (Hair et al., 2017). Accordingly, "The general

objective is to see if there are statistically significant differences between individual group models" (Hair et al., 2017, p. 42). As Hair et al. (2017) further noted,

> Technically, a multigroup analysis tests the null hypothesis H0 that the path coefficients are not significantly different…which amounts to the same as saying that the absolute difference between the path coefficients is zero…The corresponding alternative hypothesis H1 is that the path coefficients are different… (p. 283).

Accordingly, there are several tests associated with MGA, as Garson (2015) outlined:

- "1. PLS-MGA: This non-parametric significance test finds a difference to be significant if the p-value is smaller than 0.05 or larger than 0.95 for the difference of group-specific path coefficients. This method (see Henseler et al., 2009) is an extension of the original nonparametric Henseler's MGA method as described, for example, by Sarstedt et al., 2011, and is the most commonly used test.
- 2. Parametric Test: This is a similar method but is parametric, assuming that groups have equal variances.
- 3.Welch-Satterthwaite Test: This is an alternative parametric test, assuming unequal variances between groups." (p. 180).

All three can be evaluated easily to determine if there is a statistically significant difference in the structural model results. Table 7 summarizes the group comparisons.

Table 7

Summary of MGA

Group Comparison	N	DAC -> Business Value (Path Coefficient)	Difference	PLS-MGA: p-value	Parametric test: p-value	Welch-Satterthwaite: p-value
Size						
Micro	170	0.630	-0.083	0.230	0.229	0.232
Non-Micro	130	0.713				
Ownership Status						
Control Owner	219	0.686	0.046	0.551	0.545	0.540
Non-Control Owner	81	0.640				
Industry						
Product and Goods Movement Industry	105	0.732	0.077	0.224	0.269	0.224
Non-Product and Goods Movement Industry	195	0.656				
Focus						
Operational Excellence	37	0.621	-0.060	0.598	0.550	0.568
Non-Operational Excellence	263	0.681				
Customer Intimacy	44	0.570	-0.111	0.273	0.254	0.277
Non-Customer Intimacy	256	0.681				
Product and Service Leadership	13	0.740	0.071	N/A	0.500	0.500
Non-Product and Service Leadership	287	0.669				
Mixed	206	0.691	0.071	0.352	0.332	0.352
Non-Mixed	94	0.620				
Respondent Group (Time Period)						
First 100	100	0.712	0.054	0.416	0.439	0.424
Non-First 100	200	0.658				
Middle 100	100	0.697	0.036	0.59	0.625	0.596
Non-Middle 100	200	0.660				
Final 100	100	0.611	-0.087	0.272	0.234	0.276
Non-Final 100	200	0.698				

Table 7 shows that there were no statistically differences in the relationship between DAC and DABV across the group comparisons. Additionally, the product and service leadership strategy grouping only had 13 respondents in this category, so the results for this comparison are questionable at best. Importantly, there were no statistically significant differences in results when comparing the first third, second third, and last third of respondents suggesting that bias that would cause differences in groups based on when they responded during the survey recruitment period is not an issue.

Importance-Performance Matrix Analysis (IPMA)

An additional analysis was performed using the base model to further assess the results for insights into managerial implications. According to Wong (2019), IPMA

> Is often used in evaluation of the performance of key business success drivers. IPMA is basically an xy-plot where the x-axis shows the "Importance" (Total Effect) of business success drivers using a scale of 0 to 1, and the y-axis shows the "Performance" of business success drivers using a scale of 0 to 100. This way researchers can identify those predecessor constructs that have a strong total effect (high importance) but low average latent variable scores (low performance) for subsequent operational improvement (p. 120).

Additionally, it allows the researcher to provide more tangible guidance on what a manager (or in this setting SMB stakeholder) should prioritize. As Ringle and Sarstedt (2016) argued, "By combining the analysis of the importance and performance dimensions, the IPMA allows for prioritizing constructs to improve a certain target construct" (p. 1880). Hair et al. (2017) noted,

> "The use of unstandardized total effects allows us to interpret the IPMA in the following way: A one-unit increase of the predecessor's performance increases the performance of the target construct by the size of the predecessor's unstandardized total effect, if everything else remains equal (ceteris paribus)" (p. 276).

Also, Hair et al. (2018) noted,

> …two additional lines divide the importance-map into four areas with importance and performance values below and above average. Generally, when analyzing the importance-performance map, constructs in the lower right area (i.e., above average

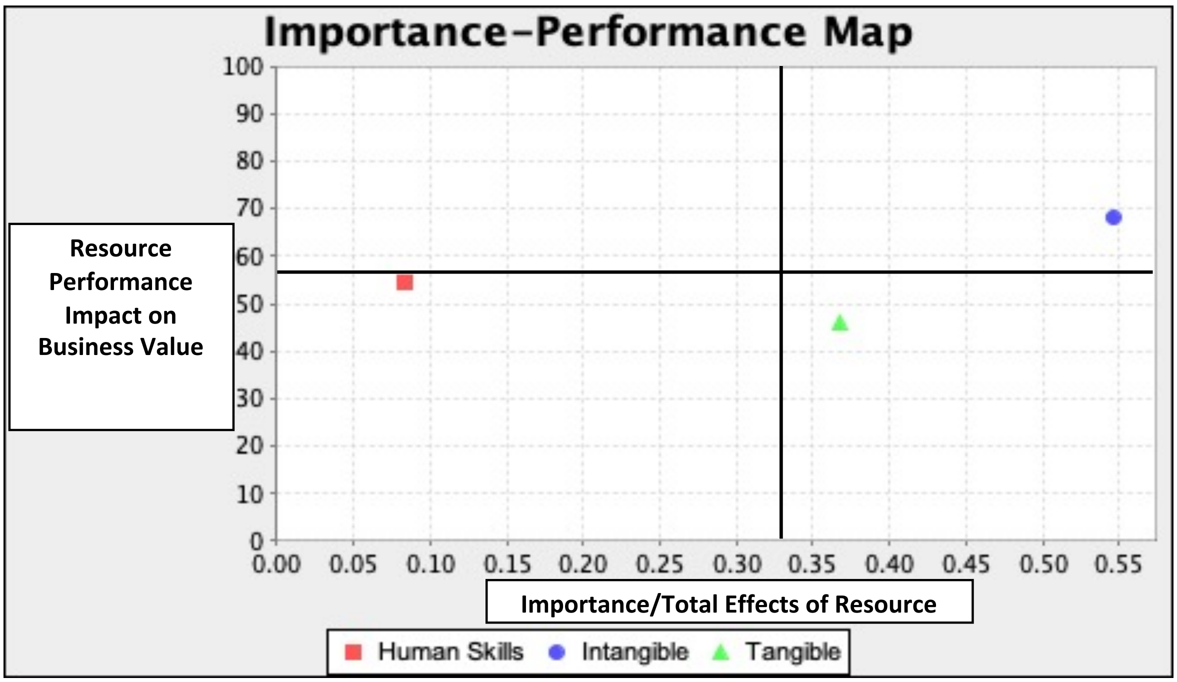
importance and below average performance) represent the greatest opportunity to achieve improvement, followed by the upper right, lower left, and finally, the upper left areas (p. 118).

In the current study, because IPMA requires that the values be on a metric scale or equidistant scale, the unstandardized latent variable scores, only available when using the IPMA procedure, were saved at each stage (Hair et al., 2018). This was to ensure that the high-order DAC constructs still fell into the Likert scale range of 1 to 7 and DABV areas followed the same scale. At the final structural model stage, the IPMA procedure was run to create an Importance-Performance Map. Using the average values for importance and performance of the three predecessor constructs (tangible, human skills, and intangible), the map was split into four quadrants. The map showing the second-order constructs and unstandardized effects is shown in Figure 17.

Figure 17

IPMA for DAC and DA Business Value at the Second-Order DAC Level

Construct	MV Performances	Construct	Business Value
Human Skills	54.52	Human Skills	0.08
Intangible	68.12	Intangible	0.55
Tangible	45.83	Tangible	0.37
Average	56.16	**Average**	0.33



As shown in Figure 17, the management priority should be given first to the development of tangible resources, followed by intangible resources, then human skills, in that order. For interpretation purposes, a one-unit increase in tangible resources, from 45.83 to 46.83, would increase the mean or average performance of DABV by 0.37 points.

Qualitative Data Analysis

Five short interviews were conducted. The primary purpose of both the quantitative and qualitative research presented in this dissertation is to impact business practice. Thus, the interview questions and summaries below provide context for the results of the quantitative

analysis through stories about real business settings. The selection of five respondents from five unique firms for interviews was purposeful. Thus, I argue that to impact practice and make the survey findings insightful for SMB practitioners, descriptions of SMB practices related to DAC and DABV will help reveal how the selected firms have allocated business resources around data analytics and realized business value.

There are different analytic techniques that can be applied qualitative interviews. For example, Mikalef et al. (2019) reported in review of three case studies developed from semi-structured interviews,

> Thus, the goal of the case studies was to further explore the interdependencies of core big data analytic capability resources and uncover emerging themes. These elements that emerged through the case studies were grouped into: (1) big data analytics strategy, (2) organizational inertia, and (3) ethics and legislation (p. 269).

This approach consisted of identifying and summarizing elements that are common or prevalent across organizations and categorizing or describing themes within the context of each case, after analyzing the individual cases (Mikalef et al., 2019).

The focus in this dissertation was on presenting context about each organization separately and to divulge how an organization might mobilize resources and institute management practices to develop DAC and realize BV. With the objective of providing more clear detail and context on individualized circumstances and unique business models, I summarized the short interviews at the organizational level first, followed by a summary discussion of general patterns.

Although these are described as vignettes as to not over reach with respect to rigor, even with the more empirical, case study approach, Symon and Casell (2012) acknowledged, “Given

the variety of forms that case studies can take and the range of sources from which data can be drawn, the development of standardized analytic techniques would be challenging" (p. 362). Additionally, while there are suggested analytic techniques for case study data, Symon and Cassell (2012) argued, "These techniques are neither rigid nor exclusive. They can be used and developed in different ways, depending on the setting…" (p. 363). Thus, the following five vignettes provide clarity on how a business owner and managers might develop BAC and realize DABV across different industries and within organizations of different size. To maintain the confidentiality of the respondents, pseudo names are used. Additionally, any unique, proprietary practices that might divulge the identity of the organization were excluded from summaries.

Vignette #1: Marketing Consulting: Marketing Strategy (MS) – 10 Employees

Company MS is a full-service, boutique marketing agency that performs brand, marketing, and sales strategy for clients including the execution of online and offline marketing campaigns. More specifically, as the owner described,

> We do digital media, we create content, we run paid advertising. We also do organic search and organic content online. So anything that you can think of that sort of runs through sort of an agency side [Company MS] performs those services. The clients that we work with are across industry verticals. So we are industry agnostic.

Additionally, MS has an approximately equal mix of business-to-consumer (B2C) and business-to-business (B2B) clients and a mix with respect to company size and maturity including Fortune 1000, midsize companies, and a small percentage of early-stage start-ups.

Internally, through a project management system, the company can collect information on upcoming projects and adjust future staffing requirements, "So if we need to hire somebody or we need to bring in a new employee, we're able to actually take a look at our production

schedule and be able to sort of forecast what our needs might be three months down the road or six months down the road." With respect to project processes, there are specific, predefined steps developed by the owner and operations manager based on the type of project and engagement that assists with everything from resources required, to strategy and execution. It is clear that the company emphasizes processes as the owner stated,

> I would say as an agency, when employees join the company, it's very apparent that we are a process driven company. We're still very creative. We do a ton of creative work, but we're very process driven and that provides a lot of stability. And so I will say that, you know, we've never heard from an employee. Gosh, I just don't know what to do or I don't know how to do this or gosh, there's no roadmap for me.

Additionally, with respect to value from internal data the owner noted, "So I would say that having sort of using the data in our project management system allows us to, again, empower our employees to be successful in their role. On the client facing side, the data absolutely helps us drive value for our clients … you can't fake the data."

The company relies on several software applications to assist with marketing activities for the clients as the owner explained:

> We're spending a significant amount of money on various platforms to collect data. So, we utilize a social media platform that allows us to track and monitor any and all engagement, impressions, metrics for our social media clients. We also use an influencer platform that allows us to track similar metrics, but then also track conversion to sales. So we're actually able to tie it directly to our clients ecommerce website so we can actually say to a client, hey, because of this one post, we reached a million people, but we also sold, you know, $20,000 in new revenue just from one post. And, we also use standard

analytics for websites like Google Analytics…I would say we probably have five to six different platforms where we track data.

In terms of the data itself, one of the challenges is the source and variety of the data, as the owner noted, “I think one of the reasons why we need five to six different platforms is because there is not one platform that allows us to analyze the data the way that we need to analyze the data.” After the data is pulled from different systems, staff can put together a summary report. As a result, the owner sees a direct correlation between having data available, the analytics that are reported by the software platforms MS has invested in, and value provided to clients:

> “…on the client side, our clients find it valuable because we're not just giving them a report with a bunch of data, we're giving them a report about why it's important to their business and how we can improve it to grow their business. And our performance with our clients based on the data does show that we are able to increase sales and revenue.”

In terms of staffing skillsets, the emphasis is on having staff who are adept at interpretation of data and metrics, “… we definitely have folks that that have 10 years of experience in the business. They're able to interpret the data just based on what we know to be true about marketing. But we don't have on staff like one person whose sole role is to be, you know, data and analytics.” Additionally, there is a continuing emphasis on software and technologies investments, as the owner stated, “So, yeah, I think our biggest challenges on the internal side, making sure that we're constantly refining our process to make it easier. Right. But then on the client side, we're always looking for the right technology to support our clients’ growth.” Yet, a major factor that must be considered by a smaller agency is pricing as the owner noted,

> So there is still this big swing in these software pricing strategies that these companies have. They always try to get you for more, but then if you try to leave, they'll basically

> drop it to next to nothing. So that's a challenge for us. Honestly, on a day-to-day standpoint, is finding the right solution and then getting it for the right price. But I think in terms of what will facilitate or how will data and analytics facilitate growth? I mean, it is absolutely the reason why we are seeing a growth in our business clients that come to us to launch a product.

In summary, there is a significant emphasis on relying on data for decision-making, "Although I will tell you, in some situations you do have to do a quick gut check. But for the most part, data is what informs changes or policy changes or updates or things like that."

Vignette #2: Aerospace and Energy: Technology and Software Development (AE) – 4 Employees

Company AE develops software and technologies for the aerospace and energy sectors. This includes the development of new hardware products and software applications for various uses in different platforms including C++, MATLAB, Python, and Apple.

While not specifically focused on analyzing internal or external business data for operations, the business uses a vast amount of data related to the environment including energy data, electronics data, and images from NASA spacecrafts. As the owner noted, "we're an engineering computer company, so everything is data-driven. That's why when we talk about business data, we don't consider it much of a big deal because we are always analyzing data."

With respect to human resources, the company has an engineering, scientist staff with a highly technical background and expertise in the sciences. Using the data collected, AE staff can develop products on a project basis that are then sold. Interestingly, for the most part the data that is searched for, collected, and analyzed is publicly available online. As such, data is a major driver of the business model as the owner noted, "the data is what allows us to produce products

that are relevant to the market". Similarly with respect to challenges, a main one can be the available data specific for the application or project as the owner stated, "Well, yeah, the problem is sometimes the data is simply not available. I mean, even NASA data can be very, very hard to get".

Finally, to the extent that the company is using different types of data including unstructured data or big data, a source might include thousands of images from a spacecraft. While this might be challenging for an organization without the right resources, AE is able to handle and analyze this type of data and turn it into products with skilled engineers, scientists, and programming software.

Vignette #3: Media Communications: Media Agency (MA) – 200 Employees

Company MA provides full-service media communications services. The respondent in this instance was a manager or executive with no ownership interest the business. The company serves various clients including those in retail and shopping where the objective might be to acquire new customers for a company such that might facilitate the purchase of a good or product. As such, the team relies on data collected from the client such as that from their website and works with web partners such as Google and Facebook to collect performance information. As a result, this requires collation of data from different data sources.

MA has a data warehouse that provides "an entire infrastructure that we've built to house all of the data across all of our tiers and campaigns for years." Additionally, the team relies on a data visualization and business intelligence tools to summarize and analyze data as noted by the executive:

> Right now what we're using is a software platform called Power BI … that essentially funnels all of our data into this very easy to use dashboard; so, you can toggle partners,

> you can toggle date ranges, you can pretty much customize this data or customize the view to pull in whatever data you want, provided you have it obviously. Power BI is a really big tool that everyone on my team uses.

These BI reports can be shared both internally and externally with clients.

With respect to staff and talent, MA has an analytics team that interfaces with executives, staff that interact with clients, and a backend data team and a technology team that builds out software platforms and data warehouse infrastructure. Collectively, the process to add value to clients consists of two key steps: “So that's how I guess that's like the two step process or the analytics team, you know, gathers the data and makes it available for us, and then our team goes in and takes the data and puts it either in a recommendation or in a report or something.” Further, while data is used in the development of campaigns and changes in media strategy, it also allows for MA to summarize the impact of media activities on clients’ business performance and justify the spend on their services to their customers. Additionally, as the executive noted on a recent campaign deliverable, “So I would say, like my team and the analytics team work very closely together to get the data where it needs to be and then visualize it and write those insights so that clients understand.”

With respect to challenges or areas for improvement the executive noted that one is the availability and flexibility of data: “So an example is with right now, our views are limited to a certain set of campaign or parameters, whether that's timing budget spend, impressions, and to be able to break things down to the model or to the tactic or those types of like deeper cuts, I think, would help strengthen our insights and strengthen our recommendations.” Additionally, the executive noted that for the future, limitations to staffing resources may be a challenge as it is expected that clients will want information faster, “I think it’s going to be speed. I think that at

the end of the day, we're all people and people will have to get things done. And when clients are demanding … they'll demand updates to happen in 24 hours because they need to capitalize on time or whatever. It really still comes down to a handful of people that have to make it happen." However, it was acknowledged that better technologies may help with meeting client expectations.

In conclusion, MA relies heavily on data for every decision related to their clients: "My team especially heavily relies on data for any decision that we make pretty much."

Vignette #4: Retail: Specialty Shoe Retailer (SR) – 15 Employees

Company SR is a single location shoe retailer that primarily sells men's and women's shoes from top brands such as Hoka, New Balance, and Saucony. SR primarily relies on data generated from and stored in the point-of-sale (POS) retail inventory system that integrates with the financial accounting system. For SR, the main objective is to drive sales and as a result a key metric is inventory turnover which can easily be benchmarked against competitors or the industry average and is a means for SR to review success from data analytics.

SR hired an outside third-party that specializes in retail analytics to assist with open-to-buy or to help determine how much and what they should purchase in inventory. As the owner stated, "And that's you know the bread and butter, you know that there are two things that we can really control actually one thing we really control in our entire business and that's inventory…can't control payroll, can't you know control the fixed costs… we use analytics all the time…control your inventory, control your business." Thus, SR relies on the consulting company that uses artificial intelligence (AI) to provide forecasts that enable effective merchandise planning and open-to-buy activities.

While much of the data analytics activity is coming from outside of SR, the owner did note that internally, a merchandise buyer on staff referred to as the “techno nerd”, has an analytical background allowing them to produce various data visualizations related to the business and carry other, more technical duties including running the store website which produces numerous metrics. In addition, with respect to the sales side, customer sales detail collected in the POS system allows SR to track information on every shoe sale to every customer. This allows sales staff to leverage purchase activity, granular to the transaction level, to engage with customers and develop sales leads. As the owner stated,

> We are a service organization … you can get the product anywhere … literally, you can get it online … we basically tell our associates when new product comes in … you know, look through your customer list and call these customers, go back and look, and see, you know who bought this product, and call them, get them on the phone … and it's been maybe a year, they can go pull that up and then call the customer say oh look it's been a year, we got a pair in, they are hard to get, you want me to save it for you or mail it to you?

Thus, while internally the use of data and skills required are less complex, it is clear that adoption of business-specific software has enabled SR to leverage data and realize value in different process areas.

Interestingly, SR sees the biggest process challenge being the lack of available and timely information shared by B2B vendors/suppliers including sharing access to availability of inventory as currently an estimated one-third of suppliers provide access to stock numbers.

Vignette #5: Casualty Claim Recovery: Subrogation Services (SS) – 8 Employees

Company SS provides subrogation services for Medicare and Medicaid health plans. The company acquires health claim data and relevant healthcare information to identify other liable payers and then pursues reimbursement for funds that should be paid to Medicare and Medicaid. SS can provide superior services to clients as data collected from different sources allows the company to identify opportunities or cases for subrogation, develop a repository for claims for production to legal or other parties in cases, and, as a result of cumulating and collating data across clients, perform ongoing risk analysis related to diagnosis codes and reimbursement methodologies at various levels including geographical. This allows SS to provide additional services and share advice to clients as the owner noted,

> Well, one of the ways that we do that is because we have years of repositories of data, we're able to go back and again look at data based on population groups by counties or other geographic areas and give assessments to our clients on how much revenue they can anticipate being able to recover based on prior experiences of subrogation cases in those in those areas. I think we're also able to forecast through experiential data both on the diagnosis codes and treatment plans, what the present future cost of healthcare may look like and what areas they may be having better or less than success in and their treatment models, I'll put it that way.

The data collected and analyzed is voluminous, as the owner noted, "I mean you're talking about data files that are … you know, at a minimum hundreds of thousands, if not millions of lines…" To handle such data, SS developed a proprietary software that relies on a SQL server database and software programs to produce specific analyses and reports. Additionally, with clients who use different software systems, SS has had to develop crosswalks to put it in a uniform and

analyzable state. As such, a significant advantage in comparison to competitors is believed to be the ability to automate the processing and analysis of data from different sources quickly with their in-house system as the owner stated, “I can describe it is that where many or most of our competitors take an average of three to five days to do a medical claim review and respond to a request…we are able to do that, within 15 to 20 minutes”. More importantly, SS pays special attention to data security as there are specific compliance and security standards that must be maintained in working with health care data.

From a staffing perspective, SS has a full-time IT employee and occasionally part-time staff who assist with programming and conduct analytical testing. Collectively, the primary duty of the staff is to manage the data that comes in and out of the company; however, they also perform analytics and conduct other data management activities. On the recovery side, the information that is aggregated and parsed is used by recovery specialists who need to have a deep understanding of medical and personal injury industries and medical coding. This allows them to scan through histories of claim expenses and identify codes that correspond to injury loss that may be subject to reimbursement by a third-party payer.

With respect to challenges going forward, pressures could come from government or politics-driven changes and changes to data security requirements around personally sensitive information. Importantly, the owner noted on data-driven culture, “I think, in fact, I would go further to say, this company is not going to exist if it's not [data-driven].”

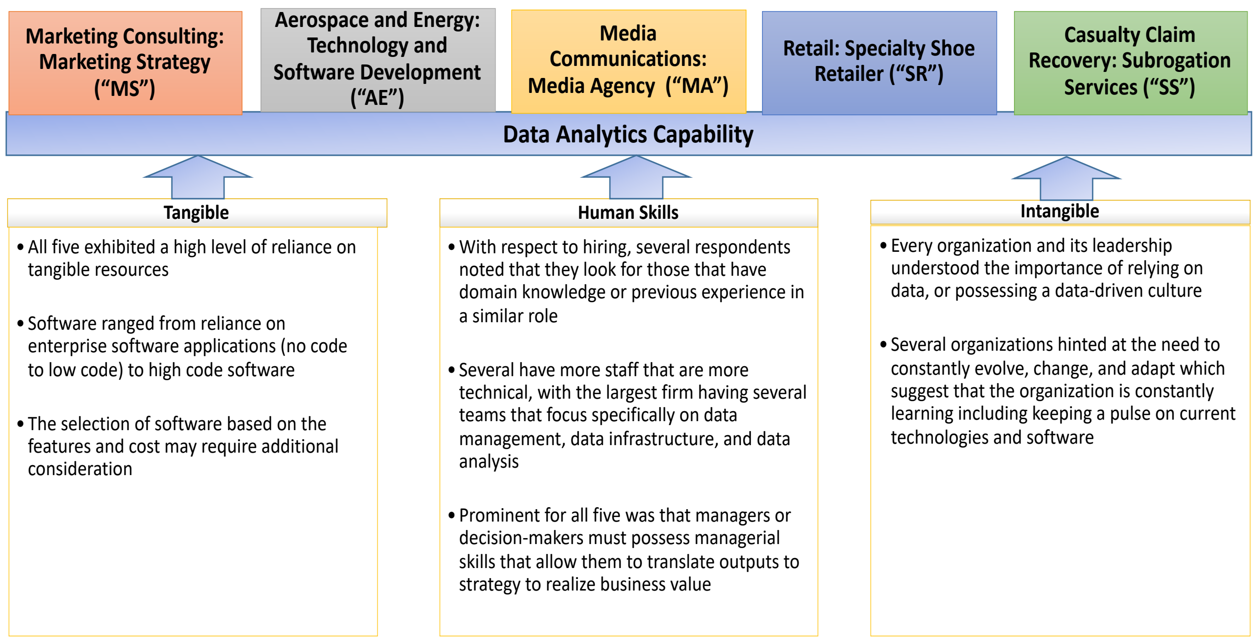
Summary of Vignette Themes in the Context of the Survey

As indicated from the literature and captured in the survey instrument, there are three main resources that can lead to the development of a data analytics capability: tangible, human, and intangible. Tangible, human, and intangible resources comprise a DAC. Further, the research

examined the relationship between DAC, the extent that big data is used, choices of software tools, and IG practices. For the interviews, the focus was more on how firms can generate business value in different areas based on resources, data, and tools utilized. Figure 18 summarizes the results.

Figure 18

Insights from Five Vignettes (Short Interviews)



Tangible Resources

All five interviewees exhibited a high level of reliance on tangible resources. While the amount and variety of data varied, all recognized the need for data, hardware, and software tools. Indeed, one respondent built a proprietary infrastructure to process and analyze data. However, the software varied which ranged from reliance on enterprise software applications (no code to low code) to high code software. Thus, the amount of customization and flexibility in data analysis varied. Yet, this is an important point: relying on analytics software that contains in-system reporting that does not require explicit coding or additional analyses can be an attractive

option to SMBs depending on the use case. Consistent with the literature, the selection of software based on the features and cost may require additional consideration in an SMB setting given a likely more limited budget in comparison to a larger firm. Further, there may be a direct connection between the need to conduct more technical analysis after data is captured by a software system as it would explicitly require the human skills and resources necessary to do so. Thus, an investment decision may require weighing the cost of software and its features or built-in capabilities and human skills or the technical skills required to manually analyze data or customize analyses to realize value.

Human Resources

With respect to human skills, all five interviewees suggested that interpretation of data or the managerial skills to be able to understand outputs from analyses is important particularly with respect to the domain or industry. Additionally, with respect to hiring, several respondents noted that they look for those that have domain knowledge or previous experience working in a similar role. Thus, context of the information generated from data and its application to specific business processes appears important for the SMB firm in its ability to extract value. However, several have more staff that are more technical, with the largest firm having several teams that focus specifically on data management, data infrastructure, and data analysis. Additionally, one respondent noted that the company relies on an outside consultant, or external resource, who has domain expertise and technical resources that allow them to use data for business decision-making. Consistent with the views presented, there are several options including relying on in-system features or built-in analytics capabilities present in purchased software (i.e., requiring less technical staff but providing less flexibility), hiring technical staff focused on data and analytics, or relying on external resources or outsourcing, such as a specialized consultant. However, what

was prominent amongst all five interviewees is that the manager or decision-maker must be able to understand and make sense of the output or possess managerial skills that allow them to translate outputs to strategy to realize business value.

Intangible Resources

Every organization and its leadership understood the importance of relying on data or possessing a data-driven culture. This was particularly prominent amongst those where the SMBs business model was largely based on data and the resulting analyses. Additionally, several organizations hinted at the need to constantly evolve, change, and adapt which suggest that the organization is constantly learning. As such, having an organizational mindset where technology is seen as important and keeping a pulse on changes in technology that can continue to improve processes allows SMBs to realize business value from data analytics. This is consistent with literature and the main motivation for capturing data-driven culture and intensity of organizational learning as it is clear that adoption of intangible resources and human resources alone is not enough for a company to see an impact performance.

Collectively, findings from the five vignettes demonstrated that the development of a DAC requires a unique blend of resources and are consistent with the survey results.

Summary of Hypotheses Tested Based on Analytical Results

- H1: DA creates business value within SMBs through positive performance impacts on key business process areas.
    - Supported, as demonstrated by the analysis of the survey data (DAC $\beta = 0.67$, $R^2 = 45.1\%$).
- H2: Tangible, human, and intangible resources are equally important in the development of a DAC.

    - Not supported, as demonstrated by the statistical analysis of the survey data (Tangible – outer loading on DAC: 0.89, $p < 0.001$), (Human Skills – outer loading on DAC: 0.81, $p < 0.001$), (Intangible Resources – outer loading on DAC: 0.89, $p < 0.001$) suggests that they are all absolutely important. However, intangible and tangible resources are of higher relative importance (Intangible – outer weight: 0.54, $p < 0.001$, Tangible – outer weight: 0.47, $p < 0.001$, and Human Skills – outer weight: 0.12, n.s.).
- H3: DAC positively effects DABV as measured by perceived value in business process areas
    - Supported, as demonstrated by the statistical analysis of the survey data, more specifically based on the positive and significant effect of DAC on Business Value (DAC $\beta = 0.67$, $R^2 = 45.1\%$). Also, discoveries from the five vignette interviews provided additional support and context around the survey results.
- H4: IG positively moderates the relationship between DAC and DABV or increases the effect of DAC.
    - Supported, as demonstrated by the statistical analysis of the survey data, more specifically based on the positive and significant effect of IG as a moderator between DAC and BV (DAC*IG $\beta = 0.11$, $R^2 = 54.0\%$).

In summary, the findings provide evidence that contribute to both academic research around data analytics capabilities as the model and resources required to develop a DAC were validated in an SMB, primarily owner-respondent setting, and knowledge for practitioners who are considering investment in resources with the objective of building a DAC.

CHAPTER 5: DISCUSSION

Overview

I had the objectives of understanding DA practices within SMBs, identifying firm resources and practices that contribute to a DAC, and to evaluate the relationship between DAC and DABV to investigate whether business value can be realized from data analytics in an SMB setting. The research goals were expressed as research questions.

A comprehensive review of the literature was conducted to develop a conceptual research model and set of hypotheses related to the research questions. A survey was developed based on instruments used in previously published peer-reviewed research, with some modifications for SMB contexts. The research model and hypotheses were tested by applying SEM to analyze responses received from 300 SMB owners and senior executives. The results provide empirical support that effective use of data analytics can impact business value in SMBs. In addition, the results provide empirical support that the resources required to form a DAC in SMBs include tangible, human, and intangible resources, and that the formation of a DAC has a positive effect on DABV. In addition, IG, a relatively untested construct in prior research, enhances the relationship between DAC and DABV. This discussion summarizes the implications for theory, business practice, limitations to the study, and recommendations for future research.

Implications for Advancing Theory

There are several theoretical implications of the research. The first implication relates to the explicit testing of the theoretical model of DAC in a new sample and target population. To my knowledge, this is the first application of the DAC model, dependent on the foundation of the resource-based view of the firm, to be tested in an exclusively SMB setting, specifically in the U.S. In a similar vein, the sample differs from previous research in that the respondents were

primarily business owners who had a financial stake in the company. This differs from previously studied groups in which the respondents were primarily executives, often in an IT or technology executive role, and additionally were recruited from an international or mixed setting. Additionally, while the technique applied here, PLS-SEM, has the potential to handle smaller sample sizes, this sample was larger by almost 100 participants or a 48.5% increase in the largest sample published to date (Mikalef et al., 2020b) using the DAC research model.

While the research model leveraged constructs developed and tested by Gupta and George (2016), the operationalization of the higher order data analytics capability construct in this dissertation focused simply on DA as opposed to ‘big data analytics.’ This dissertation validated that the resources required to form a DAC within SMBs are consistent with those required to form a BDA capability in larger firms. This is consistent with claims that the DAC model and resources identified in previous research for cultivating effective BDA are consistent with the resources required to develop BACs. I argue that based on this finding, it is likely that these constructs may be similarly reframed to capture the resources and capabilities necessary to develop business intelligence capabilities. Future researchers should take care in implementing the constructs for future data analytics research and conduct a pre-test/pilot to verify that all items are relevant and that the language is appropriately framed.

The third theoretical implication is that the DV for a data analytics outcome or target that was examined was DABV. This was operationalized as a higher order construct derived from five process areas where value can be realized by a business: customer relations, marketing and sales, product and service enhancement, production and operations, and supplier relations. Previous research has been on the impact of DAC on other business capabilities, and more generally, single order performance measures. As value can be realized in different process areas

or firm function areas, I argue that a multi-dimensional measure of DABV, acting as proxies for the impact of data analytics on business performance, provides a more comprehensive measurement in a cross-sectional study. Previous research has shown that perceptual measures of business value are strong proxies for actual business performance. Also, with the five process areas, researchers have often focused on the business value of IT (generally), however this study specifically sought to understand DABV. Indeed, the results suggest that DAC positively impacts the business value that can be realized from data analytics. A further contribution of this research is that DAC has the potential to impact a diverse set of perceptual measures of value.

The fourth theoretical implication is the measurement and testing of the contribution of IG to enhancing the realization of DABV. IG has been proposed in previous research as a moderator between DAC and value constructs (i.e., that the presence of IG is expected to enhance the value realized from DAC). However, the relationship between DAC and DABV had yet to be investigated empirically. The proposition that IG may act as a moderator was supported in this research; the moderation effect was positive and significant. While there are obvious business implications of this finding, the theoretical contribution that IG should also be considered by researchers when attempting to understand the relationship between analytics capabilities and firm performance is also clear: the presence of IG practices may act as a strong complement to DACs in enhancing the realization of business value.

A fifth implication relates to the explicit measurement of data types in the survey instrument including the capture of big, unstructured data analyzed versus traditional business data types, and the calculation of overall data scope. Additionally, the frequency of use of software tools was also captured. I argue that these may be a suitable way to collectively capture maturity level of big data use. In other words, latent constructs around tangible resources that

capture level of practices used may not be enough to understand how mature a firm may be with respect to reliance on advanced analyses or automated analyses. Indeed, not every data type should be considered equal in terms of value potential: a simple spreadsheet is different from a complex data source that may capture detailed machine or human activities. However, both may be useful for a business owner to obtain insights that impact decision-making.

Additionally, the sophistication of software that a firm may use frequently goes together with the level of analytics and types of analyses that the firm can conduct and, in turn, the potential value that can be realized. The results suggest that a higher level of big data maturity is associated with DABV realized. Again, to my knowledge, scholars have not attempted to integrate types of data, frequency of software used, or analytics maturity in studying the relationship between DACs and firm performance targets. This would contribute to the knowledgebase of big data research of capturing DAC maturity, or what some researchers might describe as value creation mechanisms between capabilities and value targets (Grover et al., 2018).

A sixth theoretical implication stems from the mechanics and technical steps in the construction of the PLS-SEM. In previously published research, authors have used the repeated-indicator approach to develop the higher order DAC construct. In this research, I used a several stage approach to avoid some of the identified issues that can emerge when using the repeated-indicator approach. This suggests that the DAC research model can be operationalized using an alternative analytical approach to develop the higher order construct.

Another research implication is demonstrated by the benefits of using qualitative research to better understand the results and the processes through which resources are used to develop a DAC. While future research may require more in-depth qualitative studies (i.e., multiple

employees interviewed, longitudinal data collection), this is the first study that focuses on the studying of the resource constructs in more detail to understand contextually how firms can realize value. At the very least, the short interviews that produced the five vignettes provide some data-driven stories around the logistics of developing a DAC. In this dissertation, the qualitative interviews revealed that the alternative paths for SMBs, which are often constrained by limited financial resources, to acquire resources may involve weighing the costs of technology resources, people, and more importantly internal investment or external investment in both. While a large firm may be less concerned with such a decision and may do a trial and error of different options, the SMB must be mindful of costs involved and may dedicate more time up front to evaluate the acquisition options.

Implications for Business Practice

With respect to business practice, the overarching takeaway is that DACs can be developed within U.S.-based SMBs, and that more sophisticated DACs lead to higher realization of DABV. Indeed, the larger implication is that data analytics practices is not limited to large firms to develop as a competitive advantage. In other words, small business owner suggestions that “we do not need that, we are too small” or “it is probably too expensive” fall short as here this dissertation has shown that effective deployment of resources specific to data analytics can positively impact different business process areas. Yet, more so what the findings suggest is that SMB owners may need to be more meticulous about acquisition of the resources underlying their data analytics capability. As the statistical analysis of the survey data demonstrated, the resources required to develop a DAC include tangible and Intangible resources. All three resource categories were important in the development of a DAC within the SMBs surveyed.

Additionally, the investment in such resources can lead to value realized from data analytics within an SMB setting, as demonstrated by the positive effect of DAC on DABV.

The first implication, like previous research, is that to form a DAC a firm requires more than just tangible resources, human resources, and organizational and cultural practices. In other words, there are important organizational and behavioral components of firm resources which include management ability to trust in data and results from analysis to make strategic decisions. Additionally, having an organizational focus on learning new and relevant knowledge within the context of the business is important. This is a significant consideration as investment in tangible and human resources are necessary but not sufficient to develop a DAC. In other words, mere adoption alone of software, hardware, analysts, and managers is not enough to realize value from data analytics. More importantly, as pointed out in the post-hoc analyses of the model results, some insights emerged about the importance of resources for SMBs to consider and prioritize.

Second, the relative importance for each resource category varies. In terms of relative importance for contribution to the formation of a DAC, intangible resources are of most importance, followed by tangible resources, followed by the human resources. What this suggests is that SMB owners should prioritize assembling intangible and tangible resources prior to developing data analytics human resources such as hiring or developing internal data analytics staff. This is further corroborated by the importance-performance analysis conducted. However, it extends the findings from understanding what leads to a DAC to identifying what resources should be the focus for SMB owners to develop to enhance the target outcome, business value.

The mapping of importance and performance of resources further demonstrates that firms should prioritize tangible and intangible resources. However, if the objective is to enhance business value, one should first focus on tangible resources as development of data resources

(i.e., capture of data and combining of sources to facilitate analysis), technology resources (i.e., software and hardware), and basic resources (i.e., providing time and funding for data analytics projects) have the highest potential to impact business value. Of second importance in impacting business value are intangible resources, including data-driven culture (e.g., management buy-in to using analytics as opposed to instinct) and organizational learning (e.g., mindfulness of new knowledge including shifts in technology). Third in line are human resources, including technical skills and managerial skills. Again, an SMB owner may then decide to further invest in tangible resources and intangible resources once the human resources has been developed to a reasonable point. It should be stated again that all three resource areas are important, however, some have higher relative importance, and some have a higher potential to influence DABV. Thus, the advice is not to ignore human resources including managerial skills or the ability to make sense of output and understand how analyses can impact strategy but instead to prioritize the acquisition of other resources first.

Additionally, in relation to both tangible and human resources, the qualitative research revealed that SMBs may decide to acquire external or internal resources. With respect to software, a SMB firm owner may then decide to use a multi-factor or multi-criterion table to evaluate software options including features and score the features according to potential business value while also considering process areas that will be impacted. For example, Tallon et al. (2020) presented an example of scoring the following criteria when comparing firm investment in cloud-based IT solutions to on-premise IT solutions: enhanced user experience, risk of user acceptance, scalability, failover scenario, level of access to information, security infrastructure, risk of storing data internally. When assessing software that will either be used for data analytics or has data analytics embedded in it (e.g., enterprise software with built-in

analytics), the SMB owner and key deciding parties may consider listing out all the desired features and pricing including assigning weights of importance to each feature, then scoring or rating each vendor to obtain an aggregate, weighted score to identify a winner. Similarly, the owner may evaluate analytics consulting options based on criteria or attributes including reputation, experience in their industry or business application needs, data management protocols, and cost of the services with potential business value realized in mind.

This findings from the study also suggest that information or data governance serves as a complement to DACs in achieving business value. While still in infancy in terms of the defined relationship between DAC and BV, IG can enhance the value realized from data analytics. SMB owners may want to start conceptualizing data as an asset that has the potential to create value and thus may decide to classify different data sources according to importance to their organization's process areas where value can be realized. In other words, there should be intentional consideration as to what data, fields, collection activities, storage, and processing activities matter the most to achieving a positive impact on firm performance, and to reduce the expenditures on resources that create limited benefits. Additionally, the SMB may want to identify responsible parties for data (i.e., clarify data ownership), increase communications amongst parties who access and analyze data, and pay attention to how data is being used.

More importantly, the significance of the research is that an SMB owner or manager may assess their analytics-capability levels using the survey instrument developed and employed in this research. As previous studies have suggested, this survey instrument may be used to evaluate the caliber of the firm's data analytics resources and capability and identify potential areas to develop. As Gupta and George (2016) noted, "The survey instrument presented in this paper can be used by organizations to determine the resources they have in abundance and the resources

they lack" (p. 13). Similarly, the DAC survey instrument can be used to assess the level of sophistication in resources acquired in different areas. More specifically, to provide a more direct comparison, the researcher can filter the respondent firms who reported achieving higher DABV so that an SMB firm owner or manager can benchmark their resource levels with the average or median resource levels of those that have realized relatively higher business value in this sample. Thus, an SMB can get a direct comparison of where they are on the resource spectrum with the target of realizing higher business value. Indeed, this may guide the realization certain tangible resources, human resources, or intangible resources that the firm lacks.

As a final point, an SMB owner or manager considering further investment in data analytics-specific resources should consider how value in such investments will be tracked. As Tallon et al. (2020) pointed out,

> A useful starting point is to trace IT impacts back from five broad categories … operations savings, performance improvements, increased sales, labour savings, and shorter conversion cycles or to critical business processes within the value chain such as supplier relations, customer relations, product and service enhancement, production and operations, and sales and marketing support (p. 9).

As a result, practitioners may consider measuring value realized prior to acquisition and periodically revisit these value scores. Similarly, the items and scales specific to the five process areas employed here can be used to evaluate the success or impact of acquisition of resources.

Limitations

There are some limitations with the research that should be addressed. First, the research conducted here is the first known application of a general DAC model as opposed to big data analytics capability model. As a result, the model and measurement items as worded should be

tested with and applied to other samples including to international firms and respondents in non-ownership roles. Additionally, while the focus for larger firms has been around big data, the model and instrument should be tested on a sample that includes both small firms and larger firms to evaluate if there are group differences. In other words, do results from larger firms when compared to SMBs suggest that SMBs have developed a different levels of data analytics capabilities and different levels of business value realized?

Second, while the target sample was SMBs, it may not be representative of the industry representation for firms with less than 500 employees in the U.S. In other words, there was limited testing on the representativeness of the firms surveyed as the researcher did not have a comprehensive list of SMBs that could be randomly selected and further stratified by characteristics that may cause differences in management practices that should be studied to better explain the phenomenon. Future researchers may be able to obtain a more comprehensive and representative list of owners and managers to solicit for participation and better control for different SMB characteristics.

Third, the survey data here was self-reported. Aside from those that participated in follow-up interviews, there was no way for me to directly verify responses aside from analyzing the statistical properties of the survey data. Thus, there may be bias present in the response data that could not be controlled for. In a similar vein, while I sought to capture perceived value, no comparisons were done to actual performance to test the correlation between perceived value realized in different process areas to tangible, financial performance measures.

Fourth, the short and semi-structured interviews were conducted with a single person. Thus, they were not in-depth and did not include discussion with additional staff which may be useful to get a more comprehensive understanding of how a DAC can be created within an SMB

to realize value. Additionally, it may be important to capture perceptions from different staff including those that conduct data analytics to evaluate if these perceptions corroborate with owners and management.

Recommendations for Future Research

There are some limitations in the research, some of which tie to recommendations for future research. The first would to be expand the sample to a larger population including, if possible, a random selection of a more comprehensive group of SMBs. Additionally, I recommend surveying both SMBs and larger firms or those with more than 500 employees to explore possible differences in DACs and value realized. As part of this recommendation, I suggest that statistical tests be conducted to evaluate the representativeness of participating firms for generalizability and extrapolation purposes.

Additionally, it is recommended that IG continue to be tested in other settings to confirm the moderating relationship with DABV. Along those lines, future research should continue to test the relevance of the newly created data types and software tools measures. Future researchers may have different perspectives on what constitutes big data maturity, which may include further consideration of the different V's (e.g., volume, velocity, veracity).

I further recommend that a longitudinal study be performed on the acquisition of resources over time and how business value evolves over time. Such a study would reveal the impact of the timeliness of acquisition and development of resources as well as the relationship between partial or full investments in resources impacts on business value realized.

Additionally, should future researchers have the resources available to collect actual performance data related to participating businesses then this should be compared with the business value areas to understand the correlation between perceived business value realized and

financial performance. While scholars have shown perceptual measures are valid proxies of business performance, I believe it is important to test the relationship between specifically DABV and business performance, given that DABV is to some extent a new construct.

Finally, researchers should continue to leverage qualitative research such as interviews and in-depth case studies to compare quantitative research results as data analytics is a fast-emerging field that is constantly evolving because of external factors. Software tools and available data sources are constantly changing. In addition, labor force training and higher education for data analytics professionals continues to shift. Further, government policies around data such as legislation around data privacy and data ownership may require additional or different resources to maintain a data analytics capability. Thus, the capture of these external factors through qualitative research may eventually develop into constructs for use in quantitative research.

Conclusion

The research presented contributes to both theory and practice. Using a survey and resulting data, I employed PLS-SEM and found a positive relationship between DAC and BV with a sample that comprised of 300 respondents from U.S.-based firms of less than 500 employees. After the data was analyzed, five case study interviews were conducted which corroborate the statistical results and provided additional context for SMBs. The findings include guidance for SMB owners and managers on the resources required to develop a DAC. Additionally, the findings provide guidance on what resources should be prioritized for building a DAC in an SMB setting. This dissertation provides a significant contribution as it demonstrates through empirical research that data analytics when employed effectively through the acquisition of certain resources can lead to the realization of business value in smaller, U.S. businesses.

As an experienced analytics professional, I see it as important to translate this research experience to five reflections around business practice. With respect to SMBs and use of data analytics:

- Firms that are low in human resources such as technical or managerial talent specific to data analytics can still realize business value from data.
- Adoption of data and technology such as hardware and software, and allocation of time and money to data analytics projects, should be complemented with the appropriate organizational practices including a data-driven culture and eagerness to learn and adapt if the intent is to produce value.
- A viable option for SMBs is to consider software with analytical capabilities embedded in it. With the sophistication of software increasing and cloud-based solutions available, out of the box solutions with the right capabilities may facilitate effective use of data for decision-making.
- An additional viable option is to hire a consultant who has expertise in both data analytics and the value area(s) that the firm seeks to target.
- Information or data governance is an emerging practice that may enhance the value realized from data analytics and from a practical perspective makes sense: If a firm and its leadership provide attention to policies, procedures, and practices around data, they are likely in a better position to realize value from data analytics than one that is not.
- Adoption and frequency of use of the popular Big Data and more advanced software tools can further enhance value however this may not be a necessity for everyone.

REFERENCES

Agrawal, K.P. (2017). Investigating Organizational Adoption of Big Data Analytics (BDA) Technology. AMCIS 2017 Proceedings.

Allen, M. (2017). The sage encyclopedia of communication research methods (Vols. 1-4). SAGE Publications, Inc

Baig, M. I., Shuib, L., & Yadegaridehkordi, E. (2019). Big data adoption: State of the art and research challenges. *Information Processing & Management*, *56*(6), 102095.

Barney, J. B., Ketchen, D. J., & Wright, M. (2011). The Future of Resource-Based Theory. *Journal of Management*, *37*(5), 1299-1315.

Brynjolfsson, E., Hitt, L. M., & Kim, H. H. (2011). Strength in numbers: How does data-driven decision making affect firm performance? *Available at SSRN 1819486*.

Chiang, R. H., Grover, V., Liang, T.-P., & Zhang, D. (2018). Special Issue: Strategic Value of Big Data and Business Analytics. *Journal of Management Information Systems*, *35(*2), 383–387.

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

Chen, L., & Nath, R. (2017). Business analytics maturity of firms: An examination of the relationships between managerial perception of IT, business analytics maturity and success. *Information Systems Management*, *35*(1), 62-77.

Coleman, S., Göb, R., Manco, G., Pievatolo, A., Tort-Martorell, X., & Reis, M. S. (2016). How Can SMEs Benefit from Big Data? Challenges and a Path Forward. *Quality and Reliability Engineering International*, *32*(6), 2151–2164.

Cragg, P., Caldeira, M., & Ward, J. (2011). Organizational information systems competences in small and medium-sized enterprises. *Information & Management*, *48(*8), 353-363.

Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches*. SAGE.

Davenport, T. H. (2009). Make Better Decisions. *Harvard Business Review*. https://hbr.org/2009/11/make-better-decisions-2

Eckerson, W. W. (2012). *Secrets of analytical leaders*. Technics Publications.

Edmondson, A. C., & Mcmanus, S. E. (2007). Methodological fit in management field research. *Academy of Management Review*, *32*(4), 1246–1264.

Everitt, B. S. (2009). *Multivariable modeling and multivariate analysis for the behavioral sciences*. CRC Press.

Fabrigar, L. R., Wegener, D. T., Maccallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, *4*(3), 272–299.

Facebook (2021). U.S. State of Small Business – 2020 – A Challenging Year for SMBs. https://about.fb.com/wp-content/uploads/2021/02/US_SSMB_Full_Report.pdf

Field, A. P. (2018). *Discovering statistics using IBM SPSS statistics: North American Edition* Fifth Edition. Sage Publications Inc.

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, *35*(2), 137–144.

Garson, G. D. (2016). *Partial Least Squares: Regression and Structural Equation Models*. Statistical Associates Publishers.

George, G., Haas, M. R., & Pentland, A. (2014). Big Data and Management. *Academy of Management Journal*, *57*(2), 321–326. https://doi.org/10.5465/amj.2014.4002

Ghasemaghaei, M. (2021). Understanding the impact of big data on firm performance: The necessity of conceptually differentiating among big data characteristics. *International Journal of Information Management*, *57*, 102055. https://doi.org/10.1016/j.ijinfomgt.2019.102055

Grover, V., Chiang, R. H., Liang, T. P., & Zhang, D. (2018). Creating strategic business value from big data analytics: A research framework. *Journal of Management Information Systems*, *35*(2), 388-423.

Grover, V., Lindberg, A., Benbasat, I., & Lyytinen, K. (2020). The Perils and Promises of Big Data Research in Information Systems. *Journal of the Association for Information Systems*, 268–293.

Gunasekaran, A., Papadopoulos, T., Dubey, R., Wamba, S. F., Childe, S. J., Hazen, B., & Akter, S. (2017). Big data and predictive analytics for supply chain and organizational performance. *Journal of Business Research*, *70*, 308-317.

Gupta, M., & George, J. F. (2016). Toward the development of a big data analytics capability. *Information & Management*, *53*(8), 1049–1064.

Hair, J. F., M., H. G. T., Ringle, C. M., & Sarstedt, M. (2017). *A primer on partial least squares structural equations modeling* (Pls-Sem). SAGE.

Hazen, B. T., Overstreet, R. E., & Cegielski, C. G. (2012). Supply chain innovation

diffusion: Going beyond adoption. *The International Journal of Logistics Management*, *23*(1), 119-134.

Headd, B. (2017, August). *The Role of Microbusiness Employers in the Economy.* https://advocacy.sba.gov/2017/08/01/the-role-of-microbusiness-employers-in-the-economy/

Hult, G. T., Hair, J. F., Proksch, D., Sarstedt, M., Pinkwart, A., & Ringle, C. M. (2018). Addressing Endogeneity in International Marketing Applications of Partial Least Squares Structural Equation Modeling. *Journal of International Marketing*, *26*(3), 1–21. https://doi.org/10.1509/jim.17.0151

Jupp, V. (2006). *The SAGE dictionary of social research methods* (Vols. 1-0). SAGE Publications, Ltd.

Kimble, C. & Milolidakis, G. (2015). Big Data and Business Intelligence: Debunking the Myths. *Global Business and Organizational Excellence*, *35*(1), 23–34. https://doi.org/10.1002/joe.21642

Kiron, D., Prentice, P.K., & Ferguson, R.B. (2012). Innovating with analytics. *MIT Sloan Management Review*, *54*(1), 47–52.

Kulkarni, U., Robles-Flores, J., & Popovič, A. (2017). Business Intelligence Capability: The Effect of Top Management and the Mediating Roles of User Participation and Analytical Decision Making Orientation. *Journal of the Association for Information Systems*, *18*(7), 516–541.

Lavrakas, P. J. (2008). *Encyclopedia of survey research methods* (Vols. 1-0). Sage Publications, Inc.

Lee, I. (2017). Big data: Dimensions, evolution, impacts, and challenges. *Business Horizons*, *60*(3), 293–303.

Llave, M. R. (2019). A Review of Business Intelligence and Analytics in Small and Medium-Sized Enterprises. *International Journal of Business Intelligence Research*, *10*(1), 19-41.

Matthews, L., Hair, J.F., & Matthews, R. (2018). PLS-SEM: the holy grail for advanced Analysis. *The Marketing Management Journal*, *28*(1), 1-13.

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

Miles, M. B., Huberman, A. M., & Saldaña Johnny. (2014). *Qualitative data analysis: A methods sourcebook*. Sage Publications.

Mooney, J.G. (1996). The productivity and business value impacts of information technology: economic and organizational analyses. University of California, Irvine Ph.D. dissertation. ISBN 978-0-591-11233-7. Downloaded from ProQuest Dissertations & Theses Global (304226480).

Mooney, J. G., Gurbaxani, V., & Kraemer, K. L. (1996). A process oriented framework for assessing the business value of information technology. *ACM SIGMIS Database*, *27*(2), 68–81.

Mikalef, P., Boura, M., Lekakos, G., & Krogstie, J. (2019). Big data analytics and firm performance: Findings from a mixed-method approach. *Journal of Business Research*, *98*, 261-276.

Mikalef, P., Boura, M., Lekakos, G., & Krogstie, J. (2020c). The role of information governance in big data analytics driven innovation. *Information & Management*, *57*(7), 103361. https://doi.org/10.1016/j.im.2020.103361

Mikalef, P., Ilias, P. O., Giannakos, M., Krogstie, J., & Lekakos, G. (2016). Big data and strategy: a research framework. In *Mediterranean Conference on Information Systems (MCIS)*. Association For Information Systems.

Mikalef, P., Krogstie, J., Pappas, I. O., & Pavlou, P. (2020b). Exploring the relationship between big data analytics capability and competitive performance: The mediating roles of dynamic and operational capabilities. *Information & Management*, *57*(2), 103169. https://doi.org/10.1016/j.im.2019.05.004

Mikalef, P., Pappas, I. O., Krogstie, J., & Giannakos, M. (2017). Big data analytics capabilities: A systematic literature review and research agenda. *Information Systems and E-Business Management*, *16*(3), 547-578.

Mikalef, P., Pappas, I. O., Krogstie, J., & Pavlou, P. A. (2020a). Big data and business analytics: A research agenda for realizing business value. *Information & Management*, *57*(1), 103237. https://doi.org/10.1016/j.im.2019.103237

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

Müller, O., Fay, M., & Brocke, J. V. (2018). The Effect of Big Data and Analytics on Firm Performance: An Econometric Analysis Considering Industry Characteristics. *Journal of Management Information Systems*, *35*(2), 488-509.

National Institute of Standards and Technology (2015, September). *Big Data at NIST*. NIST Big Data. https://bigdatawg.nist.gov/home.php

Puklavec, B., Oliveira, T., & Popovič, A. (2018). Understanding the determinants of business

intelligence system adoption stages. *Industrial Management & Data Systems*, *118*(1), 236-261.

Ramayah, T., Cheah, J., Chuah, F., Ting, H., & Memon, M. A. (2018). *Partial least squares structural equation modeling (PLS-SEM) using SmartPLS 3.0: An updated guide and practical guide to statistical analysis* (2nd ed.). Pearson.

Ringle, C. M., & Sarstedt, M. (2016). Gain more insight from your PLS-SEM results. *Industrial Management & Data Systems*, *116*(9), 1865–1886. https://doi.org/10.1108/imds-10-2015-0449

Ruel, E. E., Wagner, W. E., & Gillespie, B. J. (2016). *The practice of survey research: theory and applications*. SAGE.

Sarstedt, M., Hair, J. F., Cheah, J.-H., Becker, J.-M., & Ringle, C. M. (2019). How to Specify, Estimate, and Validate Higher-Order Constructs in PLS-SEM. *Australasian Marketing Journal*, *27*(3), 197–211. https://doi.org/10.1016/j.ausmj.2019.05.003

Shah, S., Soriano, C.B., & Coutroubis A.D. (2017). Is big data for everyone? the challenges of big data adoption in SMEs. 2017 IEEE International Conference on Industrial Engineering and Engineering Management, pp. 803-807.

Sharda, R., Delen, D., & Turban, E. (2020). *Analytics, data science, et artificial intelligence: systems for decision support*. Pearson.

Singleton, R., & Straits, B. C. (2018). *Approaches to social research*. Oxford University Press.

Tallon, P. P. (2007). A Process-Oriented Perspective on the Alignment of Information Technology and Business Strategy. *Journal of Management Information Systems*, *24*(3), 227-268.

Tallon, P. P. (2013). Corporate Governance of Big Data: Perspectives on Value, Risk, and Cost. *Computer*, *46*(6), 32-38.

Tallon, P. P., & Kraemer, K. L. (2007). Fact or Fiction? A Sensemaking Perspective on the Reality Behind Executives Perceptions of IT Business Value. *Journal of Management Information Systems*, *24*(1), 13–54.

Tallon, P. P., Kraemer, K. L., & Gurbaxani, V. (2000). Executives' Perceptions of the Business Value of Information Technology: A Process-Oriented Approach. *Journal of Management Information Systems*, *16*(4), 145–173.

Tallon, P. P., Mooney, J. G., & Duddek, M. (2020). Measuring the business value of it. *Measuring the Business Value of Cloud Computing*, 1–17. https://doi.org/10.1007/978-3-030-43198-3

Tallon, P. P., Ramirez, R. V., & Short, J. E. (2013). The information artifact in it governance: Toward a theory of information governance. *Journal of Management Information Systems*, *30*(3), 141–178. https://doi.org/10.2753/mis0742-1222300306

U.S. Census Bureau (2017). *2017 SUSB Annual Data Tables by Establishment Industry*. https://www.census.gov/data/tables/2017/econ/susb/2017-susb-annual.html.

van Riel, A. C. R., Henseler, J., Kemény, I., & Sasovova, Z. (2017). Estimating hierarchical constructs using consistent partial least squares. I*ndustrial Management & Data Systems*, *117*(3), 459–477. https://doi.org/10.1108/imds-07-2016-0286

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, *27*(3), 425-478.

Venkatesh, V., Brown, S. A., & Bala, H. (2013). Bridging the Qualitative-Quantitative Divide: Guidelines for Conducting Mixed Methods Research in Information Systems. *MIS Quarterly*, *37*(1), 21–54.

Wamba, S. F., Gunasekaran, A., Akter, S., Ren, S. J.-F., Dubey, R., & Childe, S. J. (2017). Big data analytics and firm performance: Effects of dynamic capabilities. *Journal of Business Research*, *70*, 356–365.

Wetzels, M., Odekerken-Schröder, G., & Van Oppen, C. (2009). Using PLS Path Modeling for Assessing Hierarchical Construct Models: Guidelines and Empirical Illustration. *MIS Quarterly*, *33*(1), 177-195.

Wong, K. K. (2019). Mastering partial least squares structural equation modeling (Pls-Sem) with SmartPLS in 38 hours. iUniverse.

Wixom, B., Yen, B., & Relich, M. (2013). Maximizing Value from Business Analytics, *MIS Quarterly Executive*, *12*(2).

Yin, R. K. (2018). *Case study research and applications: Design and methods*. Sage.

TABLES

Table 2

Descriptive Statistics of the Sample and Respondents

Factors	Sample	Proportion (%)
Overall	300	100.0%
Firm Size		
Greater than 10 People	130	43.3%
< 10 People	170	56.7%
Revenue (2019)		
$0	7	2.3%
$1 - $99K	57	19.0%
$100K - $499K	81	27.0%
$500K - $999K	27	9.0%
$1M - $4.99M	66	22.0%
$5M - $9.99M	24	8.0%
$10M - $24.99M	16	5.3%
$25M - $49.99M	8	2.7%
$50M - $99.99M	4	1.3%
$100M - $499.99M	4	1.3%
$500M+	2	0.7%
Unknown/Not reported	4	1.3%
Primary Industry (NAICS)		
11 Agriculture, Forestry, Fishing and Hunting	6	2.0%
21 Mining, Quarrying, and Oil and Gas Extraction	2	0.7%
23 Construction	23	7.7%
31-33 Manufacturing	41	13.7%
42 Wholesale Trade	13	4.3%
44-45 Retail Trade	16	5.3%
48-49 Transportation and Warehousing	4	1.3%
51 Information	12	4.0%
52 Finance and Insurance	12	4.0%
53 Real Estate and Rental and Leasing	14	4.7%
54 Professional, Scientific, and Technical Services	80	26.7%

55 Management of Companies and Enterprises	2	0.7%
56 Administrative and Support and Waste Management and Remediation Services	3	1.0%
61 Educational Services	11	3.7%
62 Health Care and Social Assistance	14	4.7%
71 Arts, Entertainment, and Recreation	12	4.0%
72 Accommodation and Food Services	4	1.3%
81 Other Services (except Public Administration)	28	9.3%
Other/Unknown	3	1.0%
Respondent Ownership Status		
Control owner (>50%) who actively operates the business	219	73.0%
Control owner (>50%) who passively manages the business	6	2.0%
Shared-control owner (exactly 50%) who actively operates the business	13	4.3%
Non-control owner (<50%) who actively operates the business	35	11.7%
Non-control owner (<50%) who passively manages the business	7	2.3%
Manager or executive with no ownership interest in the business	20	6.7%

APPENDIX A: IRB APPROVAL LETTER

PEPPERDINE

Pepperdine University
24255 Pacific Coast Highway
Malibu, CA 90263
TEL: 310-506-4000

NOTICE OF APPROVAL FOR HUMAN RESEARCH

Date: November 03, 2020

Protocol Investigator Name: Alfonso Berumen

Protocol #: 20-09-1443

Project Title: EFFECTIVE USE OF BIG DATA ANALYTICS AND ITS IMPACT ON BUSINESS PERFORMANCE WITHIN SMALL-TO-MEDIUM-SIZED BUSINESSES

School: Graziadio School of Business and Management

Dear Alfonso Berumen:

Thank you for submitting your application for exempt review to Pepperdine University's Institutional Review Board (IRB). We appreciate the work you have done on your proposal. The IRB has reviewed your submitted IRB application and all ancillary materials. Upon review, the IRB has determined that the above entitled project meets the requirements for exemption under the federal regulations 45 CFR 46.101 that govern the protections of human subjects.

Your research must be conducted according to the proposal that was submitted to the IRB. If changes to the approved protocol occur, a revised protocol must be reviewed and approved by the IRB before implementation. For any proposed changes in your research protocol, please submit an amendment to the IRB. Since your study falls under exemption, there is no requirement for continuing IRB review of your project. Please be aware that changes to your protocol may prevent the research from qualifying for exemption from 45 CFR 46.101 and require submission of a new IRB application or other materials to the IRB.

A goal of the IRB is to prevent negative occurrences during any research study. However, despite the best intent, unforeseen circumstances or events may arise during the research. If an unexpected situation or adverse event happens during your investigation, please notify the IRB as soon as possible. We will ask for a complete written explanation of the event and your written response. Other actions also may be required depending on the nature of the event. Details regarding the timeframe in which adverse events must be reported to the IRB and documenting the adverse event can be found in the *Pepperdine University Protection of Human Participants in Research: Policies and Procedures Manual* at community.pepperdine.edu/irb.

Please refer to the protocol number denoted above in all communication or correspondence related to your application and this approval. Should you have additional questions or require clarification of the contents of this letter, please contact the IRB Office. On behalf of the IRB, I wish you success in this scholarly pursuit.

Sincerely,

Judy Ho, Ph.D., IRB Chair

cc: Mrs. Katy Carr, Assistant Provost for Research

APPENDIX B: INTRODUCTION TO THE STUDY

Dear Business Owner/Representative,

My name is Alfonso Berumen, and I am a doctoral student in the Graziadio Business School at Pepperdine University. For my doctoral dissertation, I am examining the practices related to how Data and Analytics have been adopted by and impact the performance of Small-to-Medium sized, privately-owned businesses. You are invited to participate in this study. This survey being conducted is part of my doctoral dissertation research that I am completing at the Graziadio Business School at Pepperdine University. If you agree to participate, you are invited to complete a survey of business owners' perspectives on issues related to leveraging data to impact business performance in small and medium sized businesses (SMBs). The survey is anticipated to take about 20-25 minutes to complete. Participation in this study is voluntary.

In return for participation in the study, a summary-level report including aggregate results from approximately 300+ SMB firms will be shared with each survey respondent so that you can benchmark your business against others in your industry. Furthermore, the results will be helpful in understanding whether performance gains can be achieved from Data Analytics within an SMB setting, and if so which practices help in achieving such potential performance gains.

Additionally, in appreciation for your time, the researchers will be making a $15 donation to the Los Angeles Regional Food Bank (https://www.lafoodbank.org/) for every fully completed and submitted survey.

Your identity as a participant will remain strictly confidential. Personal identifiable data collected will be collected and used only to arrange follow-up interviews (if you agree to participate in these) and/or to share the final benchmarking report in appreciation for you completing the survey. Personally identifiable data will be separated from the other survey

response data and will be stored in a separate file that will be linked with an anonymous numeric "respondent ID." Additionally, personally identifiable data will be stored "off-line" in password protected files stored on flash-drives that will be kept securely in a locked safe or cabinet. Furthermore, the results of the analysis of the data will be presented in a summary format; individual responses, especially personal identifiable data, will not be shared in any way. Consequently, the risks associated with your participation in the survey are minimal.

If you have questions, please contact me, Alfonso Berumen, at (323) 687-3145 or alfonso.berumen@pepperdine.edu or my dissertation research supervisor Dr. John Mooney at (949) 223-2538 or john.mooney@pepperdine.edu.

If you are willing to participate in the survey, please click on [*this link*], or cut and paste the survey URL shown below into your web browser.

Thank you for your participation,

Alfonso Berumen

Doctor of Business Administration Candidate, Pepperdine Graziadio Business School

Survey Link: Small-to-Medium sized Business Owner Survey

APPENDIX C: SURVEY

Scale and Items:

Data sources:

Scale: Likert
Never
About once a year
About once a quarter
About once a month
Weekly
Daily
Multiple times a day

Items:
DS1. External data from third party, public agencies (e.g. U.S. Census Bureau, Open Government Data)
DS2. External data from third party private/commercial providers (e.g. Nielsen, Acxiom, Lotame, Oracle Data Cloud)
DS3. Data from internal operations and private business activities
DS4. Data from interactions and transactions with suppliers
DS5. Data from interactions and transactions with customers
DS6. Data streaming from sensors, RFIDs, Internet of Things (IoT), machines and/or vehicles
DS7. Data from social media postings, online customer reviews/ratings on third party websites, and online community conversations
DS8. Data generated from wearable devices (e.g. employee biometric or movement tracking)
DS9. Website browsing and search data (e.g. cookies) from computers and/or mobile devices
DS10. Geographic location data (e.g. customer or employee location) from computers and/or mobile devices
DS11. Cross-app messaging and communication data from computers and/or mobile devices

Software tools:

Scale: Likert
Never
About once a year
About once a quarter
About once a month Weekly
Daily
Multiple times a day

Items:
DA1. Custom dashboard reports manually produced by you or an employee of your business (i.e., produced using Tableau, Microsoft Power BI, Looker, Alteryx, Tableau, Microsoft Power BI, Google Data Studio, or another application)

DA2. Generic dashboard reports that are generated automatically by a Software as a Service SaaS provider or mobile application (i.e., Salesforce or another software application your company subscribes to)
DA3. Statistical programming software used by you or an employee of yours to analyze and/or visualize data (i.e., SAS, SQL, SPSS, STATA, R, Python, or another programming language or advanced tool used for statistical analysis)
DA4. Microsoft Excel or another standard spreadsheet software used by you or an employee of yours to analyze and/or visualize data
Other Data Analytics tools and/or software not identified above:

Management Practices:

Scale: Likert
Strongly disagree
Disagree
Somewhat disagree
Neither agree nor disagree
Somewhat agree
Agree
Strongly agree

Items:
Data:
D1. As a business we have access to very large, unstructured, or real-time streaming data
D2. As a business we integrate data from multiple internal sources into a data warehouse or data mart
D3. As a business we combine external data with internal data

Technology:
T1. As a business we have adopted different data visualization tools (e.g. Tableau, Microsoft Power BI, Looker, Alteryx, Google Data Studio, or another application)
T2. As a business we have adopted cloud-based services for processing data and performing analytics (e.g., Microsoft Azure, Amazon Web Services, IBM Cognos, Domo, or other services)
T3. As a business we have adopted data analysis software which either requires a fee or is open-source/free (e.g., Microsoft Excel, Google Sheets, SAS, SQL, SPSS, STATA, R, Python, and/or Mahout)
T4. As a business we have adopted new types of database technologies to process and analyze data such as Hadoop or NoSQL

Basic Resources:
BR1. Data analytics projects within our business are adequately funded
BR2. Our data analytics projects are given enough time to achieve their objectives

Technical Skills:

TS1. As a business we hire new employees who already have the data analytics skills and experience we need
TS2. Our data analytics staff currently have the right skills to accomplish their jobs successfully
TS3. Our data analytics staff have the appropriate education to fulfill their jobs
TS4. Our data analytics staff have the appropriate work experience to accomplish their jobs successfully
TS5. Our data analytics staff are well trained

Managerial Skills:
MS1. Our data analytics managers/supervisors understand and appreciate the business needs of other functional managers, suppliers, and customers
MS2. Our data analytics managers/supervisors are able to work with functional managers, suppliers, and customers to identify opportunities that data analytics might bring to our business
MS3. Our data analytics managers/supervisors are able to work with functional managers, suppliers, and customers to implement data analytics to bring value to our business
MS4. Our data analytics managers/supervisors are able to anticipate the future business needs of functional managers, suppliers, and customers
MS5. Our data analytics managers/supervisors have a good sense of where to apply data analysis to support business decision-making
MS6. Our data analytics managers/supervisors are able to understand and evaluate the output extracted from data analysis

Data-driven culture:
DD1. As a business we consider data a tangible asset
DD2. As a business we base our decisions on data and analytics rather than on "gut" or instinct
DD3. As a business we are willing to override our own intuition when the insights from data analysis contradict our viewpoints
DD4. As a business we continuously assess and improve the business rules in response to insights extracted from data
DD5. As a business we continuously coach our employees to make decisions based on insights derived from data and data analysis

Organizational Learning:
OL1. As a business we are able to search for new and relevant knowledge
OL2. As a business we are able to acquire new and relevant knowledge
OL3. As a business we are able to assimilate relevant knowledge
OL4. As a business we are able to apply relevant knowledge
OL5. As a business we have made concerted efforts to exploit existing competencies and explore new knowledge

Structural Practices:
STR1. In our business, responsibility for data ownership, data value analysis, and data cost management is shared by data / IT specialists and business managers
STR2. In our business, senior management oversee and assess data value and data costs

Relational Practices:
RLT1. In our business, we have conversations between data / IT specialists and business managers regarding data storage costs
RLT2. In our business, we have conversations between data / IT specialists and business managers regarding data management and data use

Procedural Practices:
PCR1. In our business, we have explicit practices and policies for data retention
PCR2. In our business, we have explicit practices and policies for data backup
PCR3. In our business, we have intentional practices for monitoring user access to data
PCR4. In our business, we have explicit practices for classifying data and/or information according to value
PCR5. In our organization, we have explicit practices for monitoring costs of data versus value of data

Business Value:
Scale: 1 to 7 - "Select a score from 1 to 7, where 1 indicates Low Value Realized and 7 indicates High Value Realized."

Items:
Supplier Relations
BV1. Data analytics has helped our business gain leverage over its suppliers.
BV2. Data analytics has helped us to reduce variance in supplier lead times.
BV3. Data analytics has helped develop close relationships with suppliers.
BV4. Data analytics has helped improve monitoring of the quality of products and services from suppliers.

Production and Operations
BV5. Data analytics has helped improve production throughput or service volumes.
BV6. Data analytics has helped improve operating flexibility.
BV7. Data analytics has enhanced utilization of machinery and equipment.
BV8. Data analytics has improved the productivity of labor.
BV9. Data analytics has helped streamline business processes.

Product and Service Enhancement:
BV10. Data analytics has helped enhance the value of products and services by embedding data analytics within them.
BV11. Data analytics has decreased the cost of designing new products and services.
BV12. Data analytics has helped reduce the time-to-market for new products and services.
BV13. Data analytics has helped enhance product and service quality.
BV14. Data analytics has supported product/service innovation.

Marketing and Sales:
BV15. Data analytics has helped track market response to pricing strategies.
BV16. Data analytics has increased our ability to anticipate customer needs.

BV17. Data analytics has enabled salespeople to increase sales per customer.
BV18. Data analytics has improved accuracy of sales forecasts.
BV19. Data analytics has enabled identification of market trends.

Customer Relations:
BV20. Data analytics has enhanced our ability to provide after-sales service and support.
BV21. Data analytics has improved product/service distribution.
BV22. Data analytics has helped us increase flexibility and responsiveness to customer needs.
BV23. Data analytics has enhanced our ability to attract and retain customers.
BV24. Data analytics has enabled us to support customers during the sales process.

APPENDIX D: SURVEY ITEM SUMMARY STATISTICS

Construct and Item			**Summary Statistics**				
Data	Mean	Median	Min	Max	Standard Deviation	Excess Kurtosis	Skewness
D1	3.49	3.00	1.00	7.00	2.00	-1.28	0.24
D2	3.34	3.00	1.00	7.00	2.02	-1.29	0.34
D3	4.41	5.00	1.00	7.00	1.92	-0.98	-0.50
Technology	Mean	Median	Min	Max	Standard Deviation	Excess Kurtosis	Skewness
T1	3.18	2.00	1.00	7.00	2.09	-1.31	0.43
T2	3.64	4.00	1.00	7.00	2.18	-1.54	0.06
T3	4.46	5.00	1.00	7.00	2.22	-1.29	-0.50
T4	2.63	2.00	1.00	7.00	1.90	-0.47	0.93
Basic Resources	Mean	Median	Min	Max	Standard Deviation	Excess Kurtosis	Skewness
BR1	3.81	4.00	1.00	7.00	1.83	-1.10	-0.13
BR2	4.06	4.00	1.00	7.00	1.72	-0.74	-0.32
Technical Skills	Mean	Median	Min	Max	Standard Deviation	Excess Kurtosis	Skewness
TS1	3.49	4.00	1.00	7.00	1.80	-1.16	-0.02
TS2	4.12	4.00	1.00	7.00	1.81	-0.85	-0.38
TS3	4.22	4.00	1.00	7.00	1.84	-0.86	-0.40
TS4	4.30	4.00	1.00	7.00	1.81	-0.75	-0.42
TS5	4.16	4.00	1.00	7.00	1.81	-0.80	-0.37
Managerial Skills	Mean	Median	Min	Max	Standard Deviation	Excess Kurtosis	Skewness
MS1	4.39	4.00	1.00	7.00	1.82	-0.64	-0.47
MS2	4.32	4.00	1.00	7.00	1.75	-0.48	-0.57
MS3	4.34	4.00	1.00	7.00	1.73	-0.42	-0.56
MS4	4.19	4.00	1.00	7.00	1.68	-0.45	-0.46
MS5	4.43	4.00	1.00	7.00	1.70	-0.37	-0.57
MS6	4.49	5.00	1.00	7.00	1.72	-0.30	-0.64

Data-Driven Culture	Mean	Median	Min	Max	Standard Deviation	Excess Kurtosis	Skewness
DD1	5.07	5.00	1.00	7.00	1.71	0.32	-1.01
DD2	4.55	5.00	1.00	7.00	1.57	-0.41	-0.44
DD3	4.73	5.00	1.00	7.00	1.51	0.24	-0.74
DD4	4.75	5.00	1.00	7.00	1.55	-0.06	-0.69
DD5	4.60	5.00	1.00	7.00	1.58	-0.19	-0.52

Intensity of Organizational Learning	Mean	Median	Min	Max	Standard Deviation	Excess Kurtosis	Skewness
OL1	5.59	6.00	1.00	7.00	1.34	2.58	-1.48
OL2	5.57	6.00	1.00	7.00	1.33	2.98	-1.54
OL3	5.53	6.00	1.00	7.00	1.31	2.32	-1.34
OL4	5.58	6.00	1.00	7.00	1.25	3.35	-1.54
OL5	5.45	6.00	1.00	7.00	1.45	1.36	-1.23

Structural Practices	Mean	Median	Min	Max	Standard Deviation	Excess Kurtosis	Skewness
STR1	4.11	4.00	1.00	7.00	1.83	-0.96	-0.29
STR2	5.14	6.00	1.00	7.00	1.59	1.00	-1.23

Relational Practices	Mean	Median	Min	Max	Standard Deviation	Excess Kurtosis	Skewness
RLT1	3.96	4.00	1.00	7.00	1.89	-1.11	-0.20
RLT2	4.21	5.00	1.00	7.00	1.90	-1.01	-0.44

Procedural Practices	Mean	Median	Min	Max	Standard Deviation	Excess Kurtosis	Skewness
PCR1	4.59	5.00	1.00	7.00	1.89	-0.89	-0.48
PCR2	5.31	6.00	1.00	7.00	1.72	0.26	-1.06
PCR3	4.61	5.00	1.00	7.00	1.88	-0.81	-0.46
PCR4	4.19	4.00	1.00	7.00	1.89	-1.09	-0.23
PCR5	3.81	4.00	1.00	7.00	1.83	-1.07	-0.03

Supplier Relations	Mean	Median	Min	Max	Standard Deviation	Excess Kurtosis	Skewness
SR_1	3.23	3.00	1.00	7.00	2.06	-1.23	0.37
SR_2	2.74	2.00	1.00	7.00	1.90	-0.74	0.75
SR_3	3.22	3.00	1.00	7.00	2.02	-1.27	0.35

SR_4	3.46	3.00	1.00	7.00	2.17	-1.42	0.24

Production and Operations	Mean	Median	Min	Max	Standard Deviation	Excess Kurtosis	Skewness
PO_1	3.80	4.00	1.00	7.00	2.09	-1.36	-0.02
PO_2	3.82	4.00	1.00	7.00	2.02	-1.27	-0.03
PO_3	3.14	3.00	1.00	7.00	2.03	-1.23	0.39
PO_4	3.84	4.00	1.00	7.00	2.04	-1.29	-0.04
PO_5	4.04	4.00	1.00	7.00	2.07	-1.28	-0.18

Product and Service Enhancement	Mean	Median	Min	Max	Standard Deviation	Excess Kurtosis	Skewness
PS_1	3.33	3.00	1.00	7.00	2.12	-1.33	0.31
PS_2	2.99	3.00	1.00	7.00	1.94	-1.00	0.51
PS_3	3.06	3.00	1.00	7.00	2.02	-1.19	0.46
PS_4	4.04	4.00	1.00	7.00	2.09	-1.31	-0.23
PS_5	3.89	4.00	1.00	7.00	2.16	-1.44	-0.14

Marketing and Sales	Mean	Median	Min	Max	Standard Deviation	Excess Kurtosis	Skewness
MS_1	3.62	4.00	1.00	7.00	2.12	-1.40	0.09
MS_2	3.77	4.00	1.00	7.00	2.10	-1.41	-0.04
MS_3	3.25	3.00	1.00	7.00	2.03	-1.16	0.33
MS_4	3.56	4.00	1.00	7.00	2.11	-1.34	0.15
MS_5	3.86	4.00	1.00	7.00	2.18	-1.43	-0.10

Customer Relations	Mean	Median	Min	Max	Standard Deviation	Excess Kurtosis	Skewness
CR_1	3.58	4.00	1.00	7.00	2.03	-1.30	0.08
CR_2	3.52	4.00	1.00	7.00	2.07	-1.37	0.09
CR_3	3.93	4.00	1.00	7.00	2.09	-1.31	-0.11
CR_4	3.87	4.00	1.00	7.00	2.16	-1.43	-0.08
CR_5	3.62	4.00	1.00	7.00	2.11	-1.37	0.06

APPENDIX E: SEMI-STRUCTURED INTERVIEW GUIDELINES

Overview: Thank you for taking the time to speak with me today. My name is Alfonso Berumen and am a Doctor of Business Administration candidate at Pepperdine University. I am performing this interview as a follow-up to a survey that was completed by you as part my research study. The goal of the survey was to seek information surrounding Data Analytics practices within SMBs. More specifically to understand whether and how SMBs have been able to achieve value from data analytics. The survey response from your organization indicated that you would be willing to participate in a follow up interview to share some additional insights on how value can be achieved by leveraging Data Analytics. I am going to ask you a series of questions around the themes identified in the survey to get a better understanding of your organization's practices and how your firm is able to impact performance or realized business value through such practices. Should you have any questions or need to stop the interview please feel free to let me know. This should take less than an hour to complete.

I just want to confirm that you provided a signed consent form before we start. **[Confirmed]**

Shall we begin?

1) **Introduction**
    a. Please tell me about what your organization does. In other words, how would you describe the business model?
    b. What your role is within the organization? [avoid double barreled questions!]
    c. Would you say your organization relies on data for business decision-making?

2) **Tangible:**
    a. What types of data are held by your organization? Is the data used and analyzed?
    b. Of this data, what would you consider to be "Big Data"?
        i. *If the organization collects Big Data*: to what extent is Big Data helping or hindering your realization of value from Data Analytics? Explain.
    c. What data technologies (e.g., hardware, software) have been adopted by your organization? Would you say this software is widely/extensively used?

3) Human Skills:

a. Does the company have staff with deep data analytics skills (e.g., that possess some technical training in data analysis)?

b. Are these staff focused solely on data analytics or do they have other responsibilities?

c. What knowledge/skills/competencies do you believe are really important to contributing to your organization's ability to realize value from data analytics?

d. Do you have any formal strategy or managerial process for data analytics?

e. Do you have any formal strategy or managerial process for the USE of the insights that emerge from data analytics?

4) Intangible:

a. How do people within the organization who use and analyze data create value in the company?

   i. How do they use data analytics to influence business decision-making?

   ii. What types of decisions do they influence?

b. What are the best ways to motivate employees to engage in data analytics? Or alternatively, to rely on the insights from data analytics to influence business decisions?

c. What is the most important thing your company has done that has made your employees more "data analytics savvy"?

d. What is the most important thing your company has done that has made your employees more analytics driven in their decision making?

e. Do you believe that there is a data-driven culture in your company?

   i. What factors have contributed in enabling/hindering this culture?

   ii. What could be improved?

5) Data Analytics Value

a. In which areas of your business has data analytics benefited your organization the most? (e.g. Customer relations, marketing and sales, product and service enhancement, production and operations, or supplier relations).

b. Could you describe an example of a successful data analytics project in one of these areas?

c. From your perspective, what are one or two things that you've done that have helped you increase the value that you achieved from data analytics (maybe in the context of the project)?

d. What are some of the main obstacles you encountered in realizing value from data analytics?

6) Future Business Value of Data Analytics

a. Going forward, what do you believe will hinder the ability of the organization to realize value from data analytics (e.g., access to different types of data, technologies, people or skills, culture)? What will facilitate the ability of the organization to realize value from data analytics?

APPENDIX F: REFLECTIVE CONSTRUCTS – DISCRIMINANT VALIDITY - HTMT

Discriminant Validity HTMT Data Analytics Capability					
Construct	**Data-driven Culture**	**Managerial Skills**	**Organizational Learning**	**Technical Skills**	
Data-driven Culture					
Managerial Skills	0.67				
Organizational Learning	0.57	0.58			
Technical Skills	0.61	0.87	0.53		
Business Value					
Construct	**Customer Relations**	**Marketing and Sales**	**Product and Service Enhancement**	**Production and Operations**	**Supplier Relations**
Customer Relations					
Marketing and Sales	0.89				
Product and Service Enhancement	0.86	0.79			
Production and Operations	0.86	0.82	0.87		
Supplier Relations	0.78	0.78	0.78	0.83	
Information Governance					
Construct	**Procedural Practices**	**Relational Practices**	**Structural Practices**		
Procedural Practices					
Relational Practices	0.70				
Structural Practices	0.80	0.86			

APPENDIX G: REFLECTIVE CONSTRUCTS – DISCRIMINANT VALIDITY – CROSS-LOADINGS

Discriminant Validity
Cross-loadings
Data Analytics Capability

Item	Basic Resources	Data	Data-driven Culture	Managerial Skills	Organizational Learning	Technical Skills	Technology	Outer Loading > Cross-Loading
BR1	**0.89**	0.42	0.44	0.49	0.31	0.54	0.46	Yes
BR2	**0.97**	0.41	0.55	0.66	0.47	0.63	0.44	Yes
D1	0.35	**0.76**	0.38	0.35	0.21	0.35	0.48	Yes
D2	0.39	**0.95**	0.49	0.39	0.20	0.35	0.59	Yes
D3	0.37	**0.67**	0.48	0.37	0.25	0.36	0.42	Yes
DD1	0.49	0.42	**0.770**	0.57	0.41	0.48	0.48	Yes
DD2	0.48	0.43	**0.837**	0.46	0.39	0.44	0.43	Yes
DD3	0.34	0.30	**0.703**	0.34	0.30	0.31	0.32	Yes
DD4	0.43	0.43	**0.852**	0.53	0.52	0.48	0.46	Yes
DD5	0.42	0.49	**0.819**	0.53	0.43	0.46	0.50	Yes
MS1	0.58	0.39	0.55	**0.90**	0.51	0.80	0.46	Yes
MS2	0.57	0.43	0.57	**0.94**	0.51	0.79	0.49	Yes
MS3	0.59	0.41	0.59	**0.95**	0.49	0.77	0.49	Yes
MS4	0.55	0.38	0.56	**0.92**	0.49	0.73	0.46	Yes
MS5	0.61	0.40	0.58	**0.92**	0.53	0.77	0.47	Yes
MS6	0.63	0.40	0.60	**0.90**	0.52	0.76	0.50	Yes
OL1	0.36	0.22	0.44	0.44	**0.87**	0.42	0.29	Yes
OL2	0.40	0.19	0.44	0.49	**0.91**	0.46	0.25	Yes
OL3	0.39	0.18	0.46	0.50	**0.94**	0.46	0.26	Yes
OL4	0.37	0.16	0.43	0.47	**0.92**	0.45	0.25	Yes
OL5	0.42	0.30	0.55	0.55	**0.83**	0.47	0.35	Yes
T1	0.30	0.55	0.44	0.39	0.23	0.41	**0.79**	Yes
T2	0.36	0.44	0.43	0.38	0.19	0.37	**0.72**	Yes
T3	0.41	0.39	0.50	0.47	0.29	0.43	**0.71**	Yes
T4	0.39	0.52	0.36	0.37	0.24	0.37	**0.83**	Yes
TS1	0.44	0.38	0.45	0.52	0.29	**0.69**	0.47	Yes
TS2	0.64	0.36	0.49	0.79	0.47	**0.94**	0.46	Yes
TS3	0.61	0.37	0.50	0.79	0.48	**0.96**	0.47	Yes
TS4	0.57	0.36	0.53	0.82	0.51	**0.95**	0.46	Yes
TS5	0.61	0.40	0.53	0.82	0.51	**0.96**	0.48	Yes