2019

Analyzing Small Businesses' Adoption of Big Data Security Analytics

Henry Mathias
Walden University

Walden University

College of Management and Technology

This is to certify that the doctoral dissertation by

Henry Mathias

has been found to be complete and satisfactory in all respects,
and that any and all revisions required by
the review committee have been made.

Review Committee
Dr. David Gould, Committee Chairperson, Management Faculty
Dr. Anthony Lolas, Committee Member, Management Faculty
Dr. Thomas Butkiewicz, University Reviewer, Management Faculty

Chief Academic Officer
Eric Riedel, Ph.D.

Walden University
2019

Abstract

Analyzing Small Businesses’ Adoption of Big Data Security Analytics

by

Henry Mathias

MCA, Anna University, 1995

BSc, Bharathidasan University, 1992

Dissertation Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Philosophy

Management

Walden University

May 2019

Abstract

Despite the increased cost of data breaches due to advanced, persistent threats from malicious sources, the adoption of big data security analytics among U.S. small businesses has been slow. Anchored in a diffusion of innovation theory, the purpose of this correlational study was to examine ways to increase the adoption of big data security analytics among small businesses in the United States by examining the relationship between small business leaders’ perceptions of big data security analytics and their adoption. The research questions were developed to determine how to increase the adoption of big data security analytics, which can be measured as a function of the user’s perceived attributes of innovation represented by the independent variables: relative advantage, compatibility, complexity, observability, and trialability. The study included a cross-sectional survey distributed online to a convenience sample of 165 small businesses. Pearson correlations and multiple linear regression were used to statistically understand relationships between variables. There were no significant positive correlations between relative advantage, compatibility, and the dependent variable adoption; however, there were significant negative correlations between complexity, trialability, and the adoption. There was also a significant positive correlation between observability and the adoption. The implications for positive social change include an increase in knowledge, skill sets, and jobs for employees and increased confidentiality, integrity, and availability of systems and data for small businesses. Social benefits include improved decision making for small businesses and increased secure transactions between systems by detecting and eliminating advanced, persistent threats.

Analyzing Small Businesses’ Adoption of Big Data Security Analytics

by

Henry Mathias

MCA, Anna University, 1995

BSc, Bharathidasan University, 1992

Dissertation Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Philosophy

Management

Walden University

May 2019

Dedication

I thank the Lord Jesus Christ for the great privilege to begin and complete my doctoral dissertation. I dedicate this work first to the only immortal and invisible Christ who enabled me to pursue and achieve my academic dreams. I would also dedicate this work to my father, Mathias Moves, who was my constant inspiration and motivation in all my academic dreams and successes. Finally, I dedicate this dissertation to my wife, Christina Henry, who supported me all through these tough years of this doctoral journey. I would not have accomplished all these milestones without the constant support and encouragement from my family.

Acknowledgments

I would like to thank my Dissertation Chair, Dr. David Gould, for his outstanding guidance, continuous availability, and his constant motivation for me to succeed in this doctoral journey. His enthusiasm and genuine interest in making me successful, kept me energized and focused throughout my dissertation phase. I also would like to thank my Committee Member, Dr. Anthony Lolas, for his excellent input and guidance in the field of big data and quantitative methodologies. My special thanks to Dr. Thomas Butkiewicz, as my University Research Reviewer and for all his valuable contribution to my doctoral success. Above all, I thank my God and the Lord Jesus Christ for helping me to complete this lifelong lofty dream.

I am exceedingly grateful to everyone who has helped me on this doctoral journey without whose assistance, I would not have completed my doctoral program. Every small help and guidance has enabled me to complete this monumental and substantial task of completing the dissertation. I thank God for this wonderful journey and for angels who helped me to achieve this great milestone in life.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction to the Study

Introduction

Sophisticated and malicious cyber criminals have penetrated many traditional security defenses and remained undetected for a long time (Hathaway, 2014; Verizon Enterprise, 2018). Ponemon (2015) indicated that globally, the average cost of data breaches per organization has risen to as much as $6.75 million per incident. Cyber-attacks alone could slow the pace of technology innovation with a potential loss of $3 trillion in economic value in 2020 (Kaplan, Bailey, Rezek, O'Halloran, & Marcus, 2015). Small businesses are attacked because their security is weak, and they can be used as springboards into large enterprises (Mansfield-Devine, 2016). A recent innovation termed *big data analytics* is becoming a field of immense interest among information security professionals because of its ability to analyze large-scale data at an unprecedented speed and its efficiency in correlating security-related events (Cárdenas, Manadhata, & Rajan, 2013).

The goal of this doctoral study was to examine the extent of the adoption of big data security analytics among small businesses because this is an emerging and growing technology (see Marr, 2015). I used the perceived attributes of innovation, as described in the diffusion of innovation (DOI) theory, to analyze the adoption of big data security analytics among a sample of small businesses in the United States. Valier, McCarthy, and Aronson (2008) analyzed the adoption of open source software, while Powelson (2012) examined the adoption of cloud computing by small businesses to improve contributions to the IT and the economic sectors. However, exploration of the adoption of big data security analytics among small businesses has not been a focus for previous researchers. The results obtained in this study could help in understanding the propensity of information technology (IT) business leaders, such as senior engineers, architects, managers, directors, vice presidents, and senior executives, to use big data security analytics to detect and prevent advanced, persistent threats and to improve the confidentiality, integrity, and availability of systems and data, providing improved cybersecurity to small businesses.

I will cover the problem and the purpose of the study and review the background of big data security analytics among small businesses and their dependence on IT innovation for efficiency and competitive advantage in Chapter 1. In addition, the research questions along with the theoretical framework, nature of the study, definitions included in the study, assumptions, scope, and limitations of the study will be discussed. Finally, I will present the significance of the study to theory, practice, and social change.

Background of the Study

Each decade brings new threats to businesses. Nearly all businesses face security threats that use vulnerabilities in network infrastructure and software applications. Verizon Enterprise (2018) reported that more than 58% of all data breaches occurred in small businesses, and nearly 68% of the data breaches took months or longer to discover. In another study, Horton (2014) reported that 90% of data breaches affected small businesses. Such data breaches and mass attacks produce downtime, disruption of services, and increased cost of remediation (Mansfield-Devine, 2016). Ponemon (2016), in a survey on data breaches, indicated that the average cost of lost business due to data breach in the United States was $4.13 million. In another study, Ponemon (2015) indicated that globally, the average total per-incident cost of a data breach was $6.5 million per organization, and the average cost of a data breach per record was $204. The average cost of stolen goods from small and medium businesses reached nearly $880,000, and the average the cost of recovery of stolen goods reached close to $955,000 (Apurva, Ranakoti, Yadav, Tomer, & Roy, 2017). Over time, the cost of data breaches has also increased (Ponemon, 2016). In addition, this cybercrime is likely to worsen as an increasing number of organizations become more connected using the Internet (Brewer, 2014). Most often, breaches are expensive, and affect both the reputation and the business' bottom line (Horton, 2014). The fraud and financial losses due to data breaches are not limited to any one industry, which make cyber threats a cause of major concern for all businesses (Battersby, 2014). The number of cyber-attacks has increased, and the sophistication of methods used in cyber threats has increased also (see Ponemon, 2016). In addition, cyber criminals use small businesses as springboards into large businesses because the security of small businesses is weak (Mansfield-Devine, 2016). Such breaches of systems and data reduce the confidentiality, integrity, and availability of systems and data, which are critical to the sustainability and competitiveness of small businesses. Battersby (2014) asserted that a loss of a customer's trust could be more damaging than repairing the financial loss. Brewer (2014) noted that the organizations are facing continuous data breaches due to malware attacks and advanced, persistent threats, which can remain undetected for months or even years. Therefore, the focus of this doctoral research project was to reduce and prevent the effect of security threats by studying the adoption of big data security analytics by small businesses.

Businesses are collecting a significant volume and variety of data through logging, a process where applications record attack-related activities continuously (Cárdenas et al., 2013; Li & Oprea, 2016). Manual detection of advanced attacks through log analysis is almost impossible due to the large volume of data collected through logging (Li & Oprea, 2016). Traditional security information and event management tools are unable to handle large volumes of unstructured data; however, big data analytics tools are suited to handle large volumes of disparate data sets (Cárdenas et al., 2013). Big data security analytics combine the capabilities of big data and threat intelligence to detect and minimize the advanced, persistent threats (Marchetti, Pierazzi, Guido, & Colajanni, 2016). Predictive analytics based on big data use several techniques to analyze historical and current data to predict future outcomes (Gandomi & Haider, 2015). Big data security analytics are viewed as an emerging technological platform that can intelligently identify not only undiscovered patterns of attacks but also use predictive security analytics to thwart future attacks (Marchetti et al., 2016). Farrell (2016) asserted that big data analytics provide the ability to correlate logging events to detect security incidents. However, before the big data analytical tools are put to use, processes and skilled staff should be in place to analyze large sets of machine data (Farrell, 2016). A few companies are moving toward security analytics using big data. Big companies, such as Visa, have already built security models using big data and have found them to be of great value (Richards, 2013). Intelligence-driven security fueled by big data analytics is expected to gain more adoption (Richards, 2013).

Unlike small businesses, large organizations are usually early adopters of big data security analytics; however, current research indicated that both the small businesses and large organizations are vulnerable to cybercrimes (Marchetti et al., 2016; Verizon Enterprise, 2018). Hence, increasing the adoption of big data security analytics could provide increased security to small businesses since the threats could be detected and remediated using security intelligence and big data security analytics (Marchetti et al., 2016). The unprecedented acceleration in cyber threats and data breaches call for the faster adoption of big data security analytics by small businesses. Literature reviews have indicated that there was a slow adoption of big data security analytics, and the factors for this slow adoption were unknown (Greengard, 2014; Shackleford, 2013; Verma, 2017). This gap in the literature could be addressed by analyzing the factors that contribute to the slow adoption of big data security analytics. The focus of this doctoral research was to use the DOI theory to understand the slow adoption and acceptance of this important innovation among small businesses and to provide more insight on the need to adopt.

To detect malicious threats, cybersecurity professionals have used many traditional methods that are no longer sufficient to prevent the onslaught of advanced, persistent security threats (Marchetti et al., 2016). Big data, characterized by volume, variety, velocity, and value, originate from various resources, such as the Internet, mobile devices, social media, geospatial devices, and sensors (T. Hashem, Datta, et al., 2015). Big data security analytics can help to detect incoming threats using techniques, such as agile decision-making, dynamic detection of both known and previously unknown behaviors, and effective detection of malicious behaviors in real time using multifactor approaches (Marchetti et al., 2016).

Statistically, most cyber-attacks target businesses with fewer than 250 employees (Battersby, 2014). Research related to the adoption of big data security analytics among small businesses is scarce. This study could be one of the first steps for small businesses to better understand the adoption of big data security analytics, which could provide organizations sufficient grounds to allocate more funds for the effective use of this innovation. Acquiring and implementing knowledge of big data security analytics could protect small businesses against security threats and advanced, persistent threats (Marchetti et al., 2016). The results of this study could help small businesses to prioritize the prevention of security threats and eventually save the funds spent on security breach resolutions. The findings of this study could assist also in improving the quality of applications used to protect against intruders and malicious attackers. This study was needed to increase the adoption of big data technology among small businesses to protect them from advanced, persistent threats spawned by intruders and malicious attackers. This study was also needed to guide future researchers that might attempt to provide solutions for the observations obtained because technology is constantly improving for both the cyber-attackers and the defenders.

Problem Statement

Organizations are taking a long time to detect security breaches because they are silent, sophisticated, and escape the traditional methods of perimeter protection (Lindner

& Gaycken, 2014). Examples of cyber threats include spamming, botnets, denial of service, phishing, malware, and website threats (Gupta, Tewari, Jain, & Agrawal, 2017). Advanced malware enabled threats, such as advanced, persistent threats, are persistent and multistaged, with the goal of compromising systems and data to bring substantial damage (Ghafir et al., 2018). Such advanced, persistent threats could be detected by using big data security analytics (Marchetti et al., 2016).

The general problem was that small businesses' adoption of technology innovations was slow, making organizations susceptible to advanced, persistent threats from malicious sources, which prevented their economic growth and their ability to make social contributions (Greengard, 2014; Shackleford, 2013; Verma, 2017). The specific problem was that there was a lack of information available that was specific to the slow adoption of big data security analytics among small businesses to detect and prevent advanced, persistent threats from malicious sources. Although adoption of big data analytics was one of the top priorities of organizations, only 29% of executives reported that they were using big data for predictive analytics (Greengard, 2014). The security analytics survey results published by SysAdmin, Audit, Network, and Security Institute revealed that only 25% of the big data secure analytic solutions are used for monitoring threat events and reporting (Shackleford, 2013). Understanding adoption of big data security analytics could provide insights into the efficient use of this technology to detect security threats and to prevent advanced, persistent threats effectively. In addition, the adoption of this new technology could improve confidentiality, integrity, and availability of data among small businesses (Rassam, Maarof, & Zainal, 2017).

Purpose of the Study

The purpose of this quantitative, correlational study was to examine ways to increase the adoption of big data security analytics among small businesses in the United States by examining the relationship between small business leaders' perceptions of big data security analytics attributes and their adoption. An increase in adoption could detect and prevent advanced, persistent threats from malicious resources and improve the confidentiality, integrity, and availability of data among small businesses. Small businesses have been defined as firms having fewer than 250 or 1500 employees, depending on the type of industry (Small Business Administration, 2016). For the purpose of this research, I defined small businesses as firms having fewer than 250 employees. Using the DOI theory as the theoretical framework, I collected information from small businesses through an online survey instrument that was used by Powelson (2012) to assess the adoption of cloud computing. I measured the DOI constructs, which included independent variables, such as compatibility, complexity, observability, relative advantage, and trialability, and their relationship to the dependent variable, the adoption of big data security analytics, using the web-based survey instrument. Powelson provided me with permission to adapt the instrument for surveying the adoption of big data security analytics in this study.

Research Questions and Hypotheses

Through the development and use of the following research questions, I accomplished the objectives of this study by examining the correlation between each of the perceived attributes of big data security analytics, such as the relative advantage, compatibility, complexity, observability, and trialability, and the adoption of big data security analytics. The general research question addressing the research problem was: What is the likelihood of small businesses adopting big data security analytics? The specific research questions addressing the possible correlations between variables were:

Research Question 1: To what extent does the perceived attribute of innovation called relative advantage relate to the slow adoption of big data security analytics among small businesses to detect and prevent advanced, persistent threats from malicious sources?

The first hypothesis (*H*1) was postulated such that, during the prediffusion stage, the higher the level small business leaders perceive the relative advantage of big data security analytics, the greater their adoption of big data security analytics.

> $H_0$1: There is no correlation between relative advantage and the adoption of big data security analytics.
>
> H_a1: There is a positive correlation between relative advantage and the adoption of big data security analytics.

Research Question 2: To what extent does the perceived attribute of innovation called compatibility relate to the slow adoption of big data security analytics among small businesses to detect and prevent advanced, persistent threats from malicious sources?

The second hypothesis (*H*2) was postulated such that, during the prediffusion stage, the higher the level small business leaders perceive the compatibility of big data security analytics, the greater their adoption of big data security analytics.

$H_0$2: There is no correlation between compatibility and the adoption of big data security analytics.

H_a2: There is a positive correlation between compatibility and the adoption of big data security analytics.

Research Question 3: To what extent does the perceived attribute of innovation called complexity relate to the slow adoption of big data security analytics among small businesses to detect and prevent advanced, persistent threats from malicious sources?

The third hypothesis (*H*3) was postulated such that, during the prediffusion stage, the lower the level small business leaders perceive the complexity of big data security analytics, the greater their adoption of big data security analytics.

$H_0$3: There is no correlation between complexity and the adoption of big data security analytics.

H_a3: There is a negative correlation between complexity and the adoption of big data security analytics.

Research Question 4: To what extent does the perceived attribute of innovation called observability relate to the slow adoption of big data security analytics among small businesses to detect and prevent advanced, persistent threats from malicious sources?

The fourth hypothesis (*H*4) was postulated such that, during the prediffusion stage, the higher the level small business leaders perceive the observability of big data security analytics, the greater their adoption of big data security analytics.

$H_0$4: There is no correlation between observability and the adoption of big data security analytics.

H_a4: There is a positive correlation between observability and the adoption of big data security analytics.

Research Question 5: To what extent does the perceived attribute of innovation called trialability relate to the slow adoption of big data security analytics among small businesses to detect and prevent advanced, persistent threats from malicious sources?

The fifth hypothesis (*H*5) was postulated such that, during the prediffusion stage, the higher the level small business leaders perceive the trialability of big data security analytics, the greater their adoption of big data security analytics.

H_05: There is no correlation between trialability and the adoption of big data security analytics.

H_a5: There is a positive correlation between trialability and the adoption of big data security analytics.

Theoretical Foundation

The DOI theory founded by Rogers (2003) formed the basis for the theoretical framework of this quantitative study, and the DOI theory is similar to the technology acceptance model, which was used to study the acceptance of the technology (Samar, Ghani, & Alnaser, 2017). Rogers's findings have been used in several studies to examine the process of adoption of innovations. Recently, Rogers's DOI framework has been adopted for analyzing the adoption of new technologies, concepts, and ideas (Gayadeen & Phillips, 2014; Valier et al., 2008). Rogers's DOI research highlighted that the acceptance of technology could be studied using variables, such as perceived innovation attributes, innovation decision types, communication channels, social system characteristics, and change agent effectiveness. Gayadeen and Phillips (2014) confirmed that innovation is communicated through a method called *communicated channel*. Communication is an important factor in the DOI where the message gets passed from one person to another (Rogers, 2003). However, sometimes DOI can be manipulated through politics and bribery, which is termed *incentivizing* (Gayadeen & Phillips, 2014). Gayadeen and Phillips found that the DOI framework does not work well in a negatively influenced or biased environment.

Another related theory to the study of DOIs is the theory of reasoned action, which postulates that adoption behavior is dependent on an individual's attitude and the influence of external factors (Jamshidi & Hussin, 2016). The influence by external factors includes the influence of friends and family members (Maji & Pal, 2017). Rogers (2003) also asserted that an individual's knowledge about the innovation, perception of the innovation, and the societal factors surrounding the individual play a significant role in the individual's decision to adopt the innovation.

Innovation involves programs, ideas, practices, or objects that are considered as new by the potential adopters (Rogers, 2003). Following every innovation, there are early adopters, late adopters, and laggards who adopt later than other members (Gayadeen & Phillips, 2014). Generally, there are four main components of DOI: innovation, communication channels, time, and social systems (Gayadeen & Phillips, 2014; Rogers, 2003). These four components affect the widespread adoption of innovation (Harvey, 2016). Apart from the publicized information about the innovation that needs to be adopted, the individual's perception of attributes of technology innovation affects the rate of adoption (Rogers, 2003). The perceived attributes of innovations include relative advantage, compatibility, complexity, trialability, and observability (Rogers, 2003). Valier et al. (2008) added two more variables as perceived attributes of innovation: results demonstrable and voluntariness. I used the first five attributes of innovation developed by Rogers to examine the adoption of big data security analytics in this study.

Relative advantage is one of the essential innovation attributes that depicts the advantages brought to the society by the adoption of new and improved technology (Rogers, 2003). Relative advantage is measured by improved quality and enhanced productivity and performance (Powelson, 2012). Jamshidi and Hussin (2016) affirmed that the relative advantage has a positive influence on satisfaction, which suggests that the relative advantage could be one of the determinants of user satisfaction.

Compatibility is another essential innovation attribute for accepting new technology innovations and is defined as the degree to which innovation is consistent with the values, needs, and previous experience of adopters (Rogers, 2003). Jamshidi and Hussin (2016) identified compatibility as one of the facilitators of adoption of innovation. Compatibility ensures that the innovation is performing as intended and is fit for the task in the current environment.

The perceived degree of complexity is an important attribute of emerging innovation. Rogers (2003) defined complexity as the degree to which the new technology is difficult to comprehend and apply. The complexity of innovation can become a deterrent to potential adopters who plan to adopt the innovation (Jamshidi & Hussin, 2016). Technologies that are simple, easy to learn, and easy to use can enhance the DOI.

Trialability refers to the perceived ability to use the innovation. Rogers (2003) asserted that the trialability is the ability of the product to be tried for a short period, sometimes even on an installment plan. Almost all technology innovations provide trial versions that enable potential adopters to try and use the innovative product or service for a limited time before deciding to purchase or adopt the innovation. An innovation that can be tried reduces the risk for the potential adopters (Rogers, 2003).

Observability refers to the use, features, or benefits that are visible to others and are perceived by others as useful (Rogers, 2003). Social influence can influence observability since the suggestions of peers and friends about the usefulness of the product can increase the potential to adopt the innovation (Jamshidi & Hussin, 2016). Jamshidi and Hussin (2016) further posited that social influence affects the intention to use through relative advantage and usefulness. The trialability of big data security analytics has been made possible due to the ubiquitous nature of the cloud, and the observability of big data analytics has also been made possible through visualization tools.

While the theory of reasoned action was used to study diffusion using social influence, more researchers have adopted the usage of the technology acceptance model and recognized its value in studying the DOI (Powelson, 2012; Samar et al., 2017). Powelson (2012) further asserted that the technology acceptance model could be used to study the degree of usefulness to accept or reject technology. Based on the DOI theory, Valier et al. (2008) developed a theoretical model to examine the correlation between seven perceived attributes of innovation, including relative advantage, compatibility, complexity, results demonstrable, trialability, observability, and voluntariness and the adoption of innovation. Powelson's DOI model for technology adoption prediction has been adapted in this study to examine the adoption of the emerging big data security analytics. The perceived attributes of the big data security analytics include relative advantage, compatibility, complexity, trialability, and observability. In this study, I used the variable, adoption of big data security analytics, to assess the small business leaders' adoption of big data security analytics. I further analyzed the predictive relationship between the perceived attributes of innovation and the adoption of big data security analytics using statistical tests relevant to the correlational research design. The measure of perceived attributes of big data security analytics was used to measure the degree of adoption of big data security analytics by small businesses.

Nature of the Study

Scientific methodology is a system of explicit rules and procedures (Nachmias & Nachmias, 2008). The quantitative study is a type of scientific methodology to test objective theories by examining the relationship between variables using statistical methods (Powelson, 2012). The quantitative study also provides more research designs, such as covariation, manipulation, and control methods (Nachmias & Nachmias, 2008). In this study, I used a quantitative methodology with a cross-sectional survey research design to examine the relationship between the perceived attributes of big data security analytics (i.e., relative advantage, compatibility, complexity, observability, and trialability) and the adoption of big data security analytics among small businesses. Specifically, descriptive, multiple regression, and Pearson correlations were used to examine the correlation and the strength of the correlation (see Nachmias & Nachmias, 2008).

I did not choose a qualitative methodology because it did not meet the requirements of the research questions. While quantitative methods are deterministic, qualitative research methods are exploratory and could not explain the relationship between the variables. Rosenthal (2016) posited that unlike quantitative research, qualitative research offers insight into the why of people's engagements. Also, a researcher using qualitative research does not use any statistical procedures because qualitative research is based on the assumption that complex social phenomenon cannot be represented using isolated variables (Kaya, 2013). Mixed methods research is used in studies that require a combination of strengths of both the deductive capabilities of quantitative studies and the inductive capabilities of qualitative methods (Powelson, 2012). Because the exploratory, inductive nature of the qualitative approach was not applicable to measuring the relationship between diffusion attributes of big data security analytics and the adoption of big data security analytics, neither qualitative nor mixed methods were as applicable as the quantitative method for this study. Therefore, I chose the quantitative research method to find a correlation between the perceived attributes of innovation and adoption of big data security analytics.

Among the four research design components (i.e., comparison, manipulation, control, and generalization), the comparison is an operation required to prove that the two variables are correlated (Nachmias & Nachmias, 2008). Correlation between variables helps to determine the relationship between variables in mathematical terms (Donnelly, 2007). Specifically, correlational design helps to express relationships between variables by looking at statistical measures, such as covariance and the correlation coefficient (Field, 2013). In this study, I chose a quantitative, correlational method to identify the relationship between small businesses' perception of the attributes of innovation of big data security analytics and the adoption of big data security analytics to detect and prevent advanced, persistent threats. I used the Pearson correlation coefficient to measure the positive or negative correlation between the selected variables.

I did not use control experimental methods because there was no manipulation or control of the selected variables. Analysis of variance and analysis of covariance were not applicable for this study because there was no group comparison. The Chi-square test was used to study the association between variables. To study the relation between two or more variables and the adoption, multiple regression is the suggested statistical test, if the distribution of scores is normal, having no significant outliers (Lund & Lund, 2018). For comparison of one variable against the adoption of big data security analytics, simple linear regression was used and the distribution of scores was expected to be normal.

To study the adoption of big data security analytics among small businesses, I collected data at a specific time using multiple questions embedded in a web-based survey. Specifically, a cross-sectional survey with content and construct validity was used to gather data from the participants. Web-based surveys are cost-effective and are gaining more industrial acceptance than traditional surveys due to the efficiency of grouping questions (Liu, Loudermilk, & Simpson, 2014). The web-based surveys can be accessed easily through tools, such as e-mails, and there are more chances of participation by enhancing visual design and response formats (Maloshonok & Terentev, 2016). Web-based tools also can be designed to obtain evidence of informed consent while at the same time maintaining each person's anonymity. Privacy and anonymity in web-based surveys are improved by sending responses automatically to a centralized server (Denniston et al., 2010). Data stored can be encrypted and protected against accidental theft or misuse. The sampling method, cross-sectional survey, and the goodness of fit made this inquiry a nonexperimental quantitative design.

Definitions

I framed this study on concepts and terminologies specific to big data security analytics that might be new to the reader. A brief description of each of these technical terms is provided to facilitate better understanding.

Advanced, persistent threats: Attacks by surreptitious attackers who infiltrate the system, possibly through social engineering strategies and are difficult to detect (Puri & Dukatz, 2015). The advanced, persistent threat attack is comprised of five main phases: reconnaissance, compromise, maintaining access, lateral movement, and data exfiltration (Marchetti et al., 2016).

Availability: A characteristic of the system that ensures that the information or asset is complete to authorized entities, ready for use as and when required (Vona, 2016).

Big data: Data characterized by the four Vs: volume, velocity, variety, and veracity (T. Hashem, Datta, et al., 2015).

Big data analytics: Advanced analytic and parallel techniques to process large and diverse records including a variety of contents (Gahi, Guennoun, & Mouftah, 2016).

Big data processing: A set of tools and techniques to uncover hidden data from large structured, semistructured, and unstructured data (I. A. Hashem, Yaqoob, et al., 2015).

Big data security analytics: The advanced techniques that can analyze and correlate security-related data efficiently and at an unprecedented scale (Cárdenas et al., 2013).

Botnet: A collection of many malware attacked machines (Hoque, Bhattacharyya, & Kalita, 2015)

Confidentiality: The assurance that the information is not disclosed to unauthorized entities and is not available for unauthorized use (Vona, 2016).

Denial of service: A coordinated attack where the attacker uses compromised hosts to bring down the victim (Hoque et al., 2015)

Diffusion: The process of innovation communication to the members of social systems over time (Rogers, 2003).

Innovation: An idea, practice, or object that is perceived as novel by its adopter. The adopter could be an individual, organization, or another unit of adoption (Rogers, 2003).

Innovation-decision process: The process through which a unit of adoption goes through different stages including knowledge, persuasion, decision, implementation, and confirmation (Rogers, 2003).

Integrity: A characteristic of the system that allows trusted verification and prevents unauthorized modification by authorized entities in cloud storage (Cao, He, Guo, & Feng, 2016).

Phishing: A security attack where an attacker lures the victim to provide sensitive personal information by way of enticing e-mails, malwares, or social engineering (Gupta et al., 2017).

Security breach: A skillful penetration through the system stealing data or information in a short period (Gomzin, 2014).

Small business: A firm employing employees fewer than 250 or 1500, depending on the type of industry (Small Business Administration, 2016). For the purpose of this research, small businesses were defined as firms having fewer than 250 employees.

Spamming: A security attack where the attacker or a compromised host posts continuous messages, often using botnets (Hoque et al., 2015).

Structured data: Data that can be represented using tables and stored in traditional relational database management systems (Zhan & Tan, 2018).

Threat: A weakness or possibility of attack that could compromise information security by causing loss or damage to assets (Shostack, 2014).

Unstructured data: Data that cannot be easily represented in tables, such as photos, images, opinions, log files, e-mails, forums, newsgroups, crowd-sourcing systems, and sensor data (Zhan & Tan, 2018).

Value: The value is driven by mining huge volumes of data (T. Hashem, Datta, et al., 2015) to gain competitive advantage. The value of the data is dependent on eliciting semantics out of complex and intrinsically associated data on the Internet.

Variety: Different types of data collected from various sources, such as video, audio, image, text and social media networks, in either structured or unstructured format (T. Hashem, Datta, et al., 2015). Devices generate a different type of data, and users on the Internet generate different types of data. For effective analysis, multiple sources of data are required (Blazquez & Domenech, 2018).

Velocity: The speed at which data is generated (T. Hashem, Datta, et al., 2015). For example, for every second, there are more than 2 million e-mails that are being sent (H. Zhang et al., 2015). The number of e-mails sent in 1 minute has reached 204 million (Apurva et al., 2017).

Veracity: The measure of the accuracy of data and their potential use for analysis. It is also known as quality (Saggi & Jain, 2018).

Volume: Amount of data collected from various resources, such as devices, web, text and e-mail, and audio and video sources (T. Hashem, Datta, et al., 2015). Blazquez and Domenech (2018) confirmed that scientific experiments using sensors and simulations generally generate a large amount of data. Big data are inundated with huge volumes of heterogeneous structured, unstructured, and semistructured data.

Vulnerability: A weakness in the system due to lack of strong security primitives (Ashawa, 2018).

Assumptions

Assumptions are fundamental premises and prerequisites for conducting a scientific study (Nachmias & Nachmias, 2008). I will provide my assumptions related to this research study to enhance future research in a similar field of study. Assumptions are also considered unconfirmed facts that are believed to be true at the beginning of the study (see Nachmias & Nachmias, 2008). One of my primary assumptions was that the use of big data security analytics, which are usually used by big firms, could also be beneficial when they are implemented by small businesses to detect and thwart security threats. Since many cyber criminals use small businesses as springboards to break into large firms, the strategies used for thwarting security threats in big businesses seem applicable for small businesses (Mansfield-Devine, 2016). Hence, it was assumed that the big data security analytics are as useful for small businesses as they are for big businesses, since mass security attacks can bring disruption and downtime (see Mansfield-Devine, 2016).

Another assumption was that the survey participants from small businesses can comprehend the benefits of big data security analytics and the relative advantage of using big data security analytics for detecting advanced, persistent threats. Big data analytics is a relatively new field, and the usage of big data security analytics in detecting and eliminating advanced, persistent threats is gaining momentum because of their ability to correlate events logged across the enterprises (Farrell, 2016). I assumed that small businesses also could benefit from detecting and eliminating advanced, persistent threats.

As this study was about the adoption of innovation through the perceived attributes of innovation, I assumed that the participants were skilled in conceptualizing perceptions about the big data security analytics' innovation attributes. To enable the participants to understand new terminology, the big data security analytics' terminologies were adequately explained in the survey. I also assumed that the participants were capable of understanding the relatively new terms and were competent enough to answer the survey questions. People are generally biased toward the information that is known or relevant to them (Humphreys & Sui, 2015). For example, self-bias could occur when a person has either worked or had an experience in the big data security analytics area. Such self-reporting bias can be mitigated by asking for an honest and unbiased opinion while answering survey questions (Powelson, 2012). Other assumptions included: (a) the Internet was available to all participants, (b) the population sample was representative of the target population, (c) participants could read and understand English, and (d) participants would respond to the instrument truthfully.

Scope and Delimitations

Scope refers to the choice of goals, research questions, variables, and theoretical approaches to solving the problem (Williams, 2015). The scope and boundary of this study were circumscribed by my selection of a particular innovation, the attributes of the innovation, the methods of data collection, and the type of data analysis chosen. To improve the quality and depth of this study, I narrowed down the focus to measuring one particular technology, big data security analytics, among the myriad contemporary innovations. Although this proinnovation bias could have made my investigation myopic, increased usage among the IT industry warranted further investigation to understand the intention to adopt this innovation by small businesses. The results of this study have the potential to be generalized to a wider audience since big data security analytics are expected to be used across many industries (see Bardi, Xianwei, Shuai, & Fuhong, 2015; Basole, Braunstein, & Sun, 2015; Huang, Zhao, Wei, Wang, & Du, 2015; Qiu, Wu, Ding, Xu, & Feng, 2016).

Delimitations are the boundaries by which the study is purposefully restricted (Mligo, 2016). Delimitations can be defined also as the bounds of the scope of the study arising from the conscious exclusions and inclusions made by the researcher. The selected area of study called big data security analytics delimited this study to a specific scope in the area of IT. DOI research has gained acceptance in the IT area of research, with firms investing in IT to bring knowledge from external sources and innovate internal production processes (Trantopoulos, Krogh, Wallin, & Woerter, 2017). Furthermore, the DOI social system was delimited to small businesses in this study. My use of the different stages of innovation-decision, the variables describing the rate of adoption, and the prediffusion stage of innovation as the best time for diffusion measurement made this study delimited to the attributes of DOI theoretical framework (see Rogers, 2003). The boundaries associated with the DOI theory and perceived attributes of innovation described in Valier’s theoretical model, identified what was in and out of scope for this study (see Valier et al., 2008). The scope and the study delimitations along with the innovation of big data security data analytics to improve the security of small businesses made this a unique and distinctive study.

Limitations

Limitations are methodological weaknesses or flaws, while delimitations are the boundaries to which the study purposefully restricts itself (Mligo, 2016). I used a quantitative research method in which surveys contained close-ended questions. Close-ended surveys can lead to *monomethod bias* (Molina Azorín & Cameron, 2010). Another limitation of this study was the proinnovation bias that commonly exists in diffusion related studies (Rogers, 2003, p. 106). Proinnovation bias limits the researcher's vision to see the rejection of important innovation. The selection of big data security analytics for this diffusion research study has mitigated the proinnovation bias because big data analytics have found value in better fraud detection and in effective investigation of security-related incidents (see Richards, 2013). Furthermore, the innovation-decision process can lead to either adoption or rejection (Rogers, 2003). There is also a degree of uncertainty in the diffusion process which can be minimized by disseminating information (Rogers, 2003). To minimize the errors due to lack of understanding during the survey, I provided the participants with information about this technological innovation before answering questions. The job market growth in a particular technology is one of the indicators of technology adoption as it proves that it is beneficial to the potential adopter (Plouffe, Hulland, & Vandenbosch, 2001). In addition, Ghosh (2016) ascertained that the specialization in big data analytics is considered to be one of the mostly highly paid jobs, making this a suitable candidate for a prediffusion research study.

Another probable limitation of this study was the practical limitation of a collection of small businesses from a web participant pool that might not be representative of the entire population (see Nachmias & Nachmias, 2008). This limitation was overcome by increasing the sample size and opening the survey to all participants in the United States, which can mitigate the limitations of convenience sampling. This research also was dependent on the participants' understanding of the survey, which could have been another limitation of this study, since participants might have provided inaccurate responses if the questions were not correctly understood. This limitation was mitigated by sending the survey to potential decision makers within small businesses.

Significance of the Study

The results of this study could bring a significant contribution to the DOI theory, to the practice of big data security analytics, and to the society by providing more safety and security to small businesses. Big data are collected even without human intervention since the technology gadgets used today automatically collect data that includes behavioral patterns (Strong, 2014). By using the big data security analytics, it is possible that security threats and advanced, persistent threats can be detected, providing a significant contribution to the security of the organizations (Apurva et al., 2017). I will discuss the significance of this study to theory, practice, and social change in the following subsections.

Significance to Theory

DOI theory has enabled the community to understand the diffusion of technology using the five steps in the innovation-diffusion process: adoption, knowledge, decision, implementation, and confirmation (Rogers, 2003). The DOI theory also helps to understand the adopter categories: innovators, early adopters, early majority, late majority, and laggards (Rogers, 2003). Additionally, the DOI theory has become increasingly used in the IT sector (Valier et al., 2008). This study on big data security analytics brings a new dimension to the DOI theoretical framework because DOI was used in the area of security analytics, which is quite rare. Future researchers are likely to now use this theory to understand the diffusion of security-related innovations.

Significance to Practice

Advanced, persistent threats, which are sophisticated, human-driven threats, pose a great danger to the business continuity of IT businesses (Marchetti et al., 2016). As the number of devices connected to the Internet increases, there is potential for increased attack surface that could affect critical services provided by small businesses (Hathaway, 2014). Small businesses are often used as springboards to launch attacks against big enterprises (Mansfield-Devine, 2016). This research could begin to fill the gap in the literature on this topic by analyzing small businesses’ adoption of big data security analytics. More specifically, the results of this study reveal the correlation of small businesses’ prediffusion perceptions of big data security analytics’ innovation attributes and the intention to use big data security analytics to detect and prevent security challenges, such as advanced, persistent threats. By providing this correlation, this study fills the existing gap in the literature on the business practice of linking big data security analytics to small businesses’ security.

Significance to Social Change

Interpersonal communication channels help to influence the adoption of new technology (Rogers, 2003). Increased understanding of this technology could help to protect organizations from security breaches and advanced, persistent threats. Protection from advanced, persistent threats is becoming a necessity to reduce the risk of losing intellectual properties and to eliminate disruption of IT-enabled businesses (Kaplan et al., 2015). The outcomes of this study could bring substantial positive social change by bringing about a deeper understanding of the advanced, persistent threats and big data security analytics among small businesses' leaders. Using big data security analytics to detect and prevent advanced, persistent threats could reduce small businesses' cyber-risks by improving the confidentiality, integrity, and availability of systems and data (Aminzade, 2018). Personal, identifiable information can be protected from theft, and the integrity of data can be preserved by using big data security analytics. Corporate identity and reputation could be protected if small businesses understand more about this innovative technology and use it to protect their systems and data.

Additionally, small businesses can use these findings to improve the capitalization of IT and resources. The results obtained from this research could be used to improve employee development proficiencies in preparation for the anticipated increased use of big data security analytics among small businesses in areas, such as e-business, healthcare, science and technology, finance, digital marketing, supply-chain operations, security, and governance (see Ghosh, 2016). Local communities are expected to benefit by an increase in the job market for the jobs requiring specialization in big data security analytics, such as big data scientists (see Englmeier & Murtagh, 2017). Public companies and regulatory agencies could be better equipped with an understanding of this technology to facilitate collaboration between academic and business communities on projects related to big data security analytics.

Summary and Transition

Big data analytics provide an enormous amount of benefits because of their capability to process unstructured, semistructured, and structured data (Gahi et al., 2016). The tools of big data analytics also provide insights by addressing big data, which are difficult to process using traditional database techniques. Such tools can be applied to analyze security logs that are maintained for a long time to detect security breaches that are left undetected by traditional intrusion detection systems and intrusion prevention systems. The adoption of big data security analytics to detect traditional and advanced, persistent threats has been relatively slow (Greengard, 2014; Verma, 2017). I designed this study to determine the correlation between the perceived attributes of this important innovation and the adoption of big data security analytics. The purpose was to increase awareness and the adoption of this important innovation among small businesses, which could help them to detect and prevent advanced, persistent threats.

Chapter 2 will include an overview of small businesses, sources of big data, and the benefits of big data in data analytics. Also, Chapter 2 will comprise evidence from the literature demonstrating potential uses of big data analytics in various industry segmentations. Evidence of the slow adoption of big data analytics in identifying and preventing advanced, persistent threats appear in Chapter 2.

Chapter 2: Literature Review

Introduction

The specific problem addressed in this study was the lack of information available specific to the slow adoption of big data security analytics among small businesses to detect and prevent advanced, persistent threats from malicious sources. The adoption of big data security analytics could improve the integrity, confidentiality, and availability of data among small businesses (Rassam et al., 2017). The purpose of this quantitative correlational study was to examine ways to increase the adoption of big data security analytics among small businesses in the United States by examining the relationship between small business leaders' perceptions of big data security analytics attributes and their adoption.

Small businesses are firms, usually comprised of fewer than 250 or 1500 employees depending on the type of industry (Small Business Administration, 2016). For the purpose of this research, I defined small businesses as firms having fewer than 250 employees. A recent survey revealed that 58% of all data breaches occurred in small businesses, and 68% of breaches took months or longer to discover (Verizon Enterprise, 2018). The literature review for this study indicated that big data security analytics can be used to identify and thwart security threats including the advanced, persistent threats. However, the adoption of big data security analytics among U.S. businesses has been very slow (Greengard, 2014). Adoption of an innovation is possible only when the innovation is seen as something new or useful in the eyes of the consumer (Rogers, 2003). By studying the perceived attributes of innovation, it is possible to measure the inclination to adopt the innovation.

This literature review will be divided into seven sections. The first section will contain a review of small business owners’ perspectives and a description of current statistics, growth, global presence, and the status of small businesses. In the second section, I will delineate the importance of IT innovation, which provides a competitive edge for small businesses while they compete within their field. In the third section, big data, which are increasingly becoming useful for studying the large data, characterized by volume, velocity, and variety, will be defined. In the fourth section, I will describe the big data sources, which provide voluminous data that can be analyzed for insights and intelligence. The fifth section will include a discussion of big data analytics, which focuses on mining the existing data to find out actionable intelligence without compromising the integrity and validity of data. In the sixth section, big data security analytics, which uses the big data of the organizations to detect and thwart advanced, persistent threats, will be described. In the last section, I will discuss the ability of big data security analytics to improve the productivity and competitive advantage of small businesses. Finally, I will review previous studies that used the perceived attributes of innovation, such as compatibility, complexity, observability, relative advantage, and trialability, to study the adoption of big data security analytics. This literature review helped me to identify the gaps related to the adoption of big data security analytics among small businesses.

Literature Search Strategy

The bulk of my literature search took place in EBSCOHost’s electronic databases, such as Business Source Complete, Academic Search Complete, Education Resource, ProQuest Central, IEEE Xplore Digital Library, PsycINFO, ACM Digital Library, Science Direct, and Computers & Applied Sciences Complete. ProQuest’s Doctor of Philosophy Dissertations and Theses and Emerald Management Journals were also used for the literature search. I chose articles from peer-reviewed journals whose publication dates ranged from 2013 to 2018 for inclusion in this study. Keywords and phrases used singly or in combination included: *diffusion of innovation*, *DOI, small business IT innovation*, *innovation, small business America*, *small business Europe, small business Australia*, *big data definition*, *big data*, *big data analytics*, *big data security*, *big data security analytics*, *security threats*, *security threats small businesses, security threats small businesses*, *big data issues*, *big data challenges*, *big data obstacles*, *big data growth*, *big data concerns*, and *big data future*.

My primary search strategy was to review peer-reviewed journals and articles published primarily within the last 5 years. Other resources included Walden dissertations and popular books published in ProQuest online libraries. The literature review indicated the significance of small businesses in the United States and the vulnerabilities of small businesses to the cyber-attacks. As the information about small businesses using big data security analytics were very scarce, I extended my search to include electronic books.

Theoretical Foundation

I used the DOI theory to study the diffusion of big data security analytics among small businesses. In the DOI theory, developed and refined by Rogers (2003), diffusion was defined as the process by which an innovation is communicated through certain channels over time among members of a social system. Rogers also defined innovation as an idea, practice, or object perceived as new by an individual or other unit or adopter. For diffusion to take place, Rogers pointed out that there must be an idea or innovation to be diffused. The DOI has been used in the past to study diffusion in several fields, such as anthropology, early sociology, rural sociology, education, public health, communication, marketing, and management (Rogers, 2003). Recently, Rogers's DOI framework has been used for analyzing the adoption of new technologies, concepts, and ideas (Gayadeen & Phillips, 2014; Powelson, 2012; Valier et al., 2008). The characteristics of innovation as perceived by an individual include relative advantage, compatibility, complexity, trialability, and observability (Rogers, 2003). Innovations that are perceived by individuals to have greater relative advantage, compatibility, trialability, and observability, and less complexity can be adopted more easily and quickly than other innovations (Rogers, 2003). The rate of innovation adoption is categorized by five variables: perceived innovation attributes, innovation decision types, communication channels, social system characteristics, and change agent effectiveness (Rogers, 2003). Rogers further asserted that innovation could also be modified or enhanced by users in the process of adoption and implementation, giving birth to reinventions that could further diffuse invention rapidly and make it more sustainable.

In using Rogers’s (2003) diffusion theory in this study, I was focused on how the innovation (i.e., big data security analytics) was diffused within small businesses. New technologies are first considered and then initiated into organizations; once the new technologies are implemented, they can be considered diffused into the organization (see Rogers, 2003). Big data security analytics are a relatively recent innovation, having the potential to be applied in the field of IT security. The potential advantage of a new idea propels an individual to understand more about the innovation and eventually make a decision to adopt the innovation; Rogers called this process an innovation-decision process in which the individual performs an information seeking and information processing activity to reduce the risks associated with the adoption of an innovation. The diffusion of big data security analytics can be studied by analyzing the degree of perception of this technology by the decision makers in small businesses. This technology innovation is more likely to be adopted by individuals who perceive this technology to have greater relative advantage, compatibility, trialability, and observability, and less complexity (Rogers, 2003).

Rogers (2003) defined the following attributes that promote diffusion and adoption of any new technology:

1. Relative advantage: The degree to which an innovation is perceived to be better than the idea it supersedes.
2. Compatibility: The degree to which an innovation is perceived as consistent with the existing values, past experiences, and needs of the potential adopters.

3. Complexity: The degree to which an innovation is seen as difficult to understand and use.
4. Trialability: The degree to which an innovation can be experimented with, or implemented, in parts.
5. Observability: The degree to which the results of an innovation are visible to others (pp. 15–16).

Rogers (2003) affirmed that the perceptions of the presence of these five attributes influences the propensity to adopt the innovation. Technology diffusion can be studied similar to the diffusion of an idea or a news event (Rogers, 2003). In this study, I will use Rogers's model to explain the adoption of new technology, such as that of big data security analytics among the small businesses in the United States.

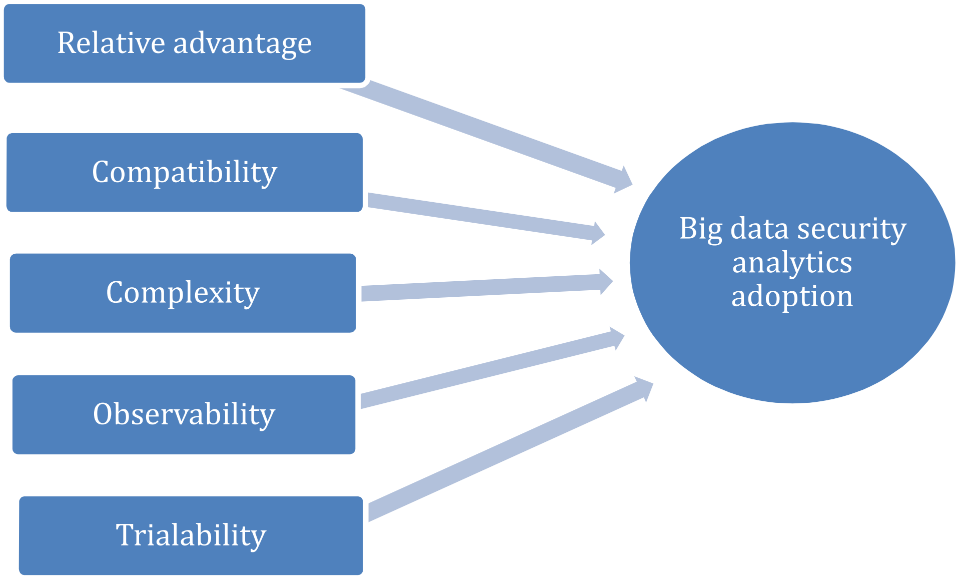


Figure 1. Perceived attributes of innovation.

Innovations are adopted only if they are considered to be new by the potential adopters (Rogers, 2003). In the DOI framework, the rate of adoption is defined as the relative speed at which an innovation is adopted by the members of a social system (Rogers, 2003). In the DOI theory, potential adopters of the innovation are categorized into five types:

1. Innovators: The people in this category are seeking new information actively and are willing to venture into innovations beyond their comfort zone to make new contacts and to learn new things (Rogers, 2003; Williams, 2015).
2. Early adopters: The people in this category tend to be more rooted in their network of relationships (Williams, 2015). The innovation needs to be accepted within their local network for them to adopt it; they tend to be more respected and viewed as normal within society (Rogers, 2003).
3. Early majority: These adopters are more thoughtful, and they take more time to adopt any innovation (Williams, 2015).
4. Late majority: The people in this category make little use of communication channels and mostly learn from their peers (Rogers, 2003). Their skepticism only allows them to adopt after they have seen the innovation work (Williams, 2015).
5. Laggards: The people in this category are the last to adopt the innovation (Williams, 2015).

The DOI theory has been used in several past research studies. Powelson (2012) used the DOI to study the adoption of cloud computing by small businesses in Arizona and observed that there was a significant correlation between compatibility, complexity, observability, relative advantage, results demonstrable, and the propensity to use cloud computing. However, there was no significant correlation between voluntariness and the propensity to use cloud computing (Powelson, 2012). Moreover, Powelson asserted that the use of DOI theory in studying the diffusion of cloud computing increased awareness among the small business leaders about the advantages of using cloud computing and created more job opportunities for those experienced in cloud computing in small businesses.

Jamshidi, Hussin, and Wan (2015) used the DOI framework to study the factors affecting the adoption of Islamic banking services in Malaysia. The results showed that the perceived attributes of innovation, including relative advantage, compatibility, complexity, trialability, and observability of Islamic banking services influenced the customer to use more features of the Islamic banking services in Malaysia. The DOI theory provided a framework to identify the factors that were more influential in affecting the decision to use or adopt Islamic banking services in Malaysia (Jamshidi et al., 2015).

DOI theory has been used in the health care industry for studying the adoption of technology innovation in hospitals. Waring and Alexander (2015) found that the DOI framework helped to gain insight into patient flow and bed management, a problem that was pervasive among healthcare organizations. The research conducted by Waring and Alexander using DOI theory produced practical suggestions regarding adoption of new patient flow and bed management systems and showed how academic research could affect healthcare organizations.

The DOI framework has been used to study adoption in the banking industry. Deb and Lomo-David (2014) used DOI theory and framework to study the factors affecting the adoption of m-banking. The study found empirical evidence of a positive relationship between perceived usefulness, perceived ease of use, and social influence, and a positive attitude towards m-banking (Deb & Lomo-David, 2014). The DOI framework helped to find a positive relationship between attitude towards m-banking and the intention to adopt m-banking. The adoption of m-banking can be increased by improving the customer's perception of benevolence, privacy, and security (Deb & Lomo-David, 2014).

Rogers's (2003) DOI theory and framework were used to study the factors that led to the acceptance and diffusion of clinical solutions in New Zealand (Nath, Hu, & Budge, 2016). Nath et al. (2016) found that both the human (clinicians) and non-human factors (the software package) influenced the adoption of the innovation. Thus, the DOI framework aided the study of the diffusion of software packages into the health care industry. The outcome of the study illuminated the agents that influenced the diffusion of the clinical software package, an IT innovation in the public health sector.

A related theory to diffusion study is the theory of reasoned action, which postulates that the adoption behavior is dependent on an individual's attitude and the influence of external factors (Jamshidi & Hussin, 2016; Maji & Pal, 2017). Rogers (2003) stated that an individual's knowledge about the innovation, perception of the innovation, and the societal factors surrounding the individual plays a significant role in the individual's decision to adopt the innovation. The four main components of DOI: (a) innovation, (b) communication channels, (c) time, and (d) social system affect the widespread adoption of innovation (Gayadeen & Phillips, 2014; Harvey, 2016; Rogers, 2003). Since DOI theory has been used in the past to study the adoption of innovations, such as cloud computing, open source, and related technologies, it makes sense to use this theory to study the adoption of big data security analytics during their diffusion stage in the IT industry. In the next section, I will provide a comprehensive literature review on small businesses, IT innovation, big data sources, big data analytics, and the adoption of big data security analytics.

Literature Review

The literature review for this study was categorized into seven topical sections. The first section will contain information about the small businesses' current statistics, global influence, and their reasons for failure. In the second section, I will highlight the influence of IT and big data innovation on small businesses. In the third section, big data, which are described as data and information that are too big to be analyzed or managed with traditional technologies will be defined (Rassam et al., 2017). In the fourth section, I will enumerate the sources of big data, which could be complex due to the nature of the data (Blazquez & Domenech, 2018). In the fifth section, I will underline the importance of big data analytics. In the sixth section, I will describe the area of focus, big data security analytics, and in the seventh section, I will describe the need for big data security analytics among small businesses, including the perspectives of small businesses and the need for big data security analytics in different parts of the globe.

Small Business Perspectives

Although small businesses in the past have provided vitality to the UK economy, the survivability of small businesses is precarious. Fewer than 65% of small businesses in the United Kingdom survived 3 years after their startup (Gray & Saunders, 2016). Even though small businesses are more likely to fail than bigger enterprises, some small businesses survived and indeed prospered (Gray & Saunders, 2016). A life cycle of a small business can be represented by four phases: formation, early growth, later growth, and stability or decline (Dodge & Robbins, 1992). Some small businesses failed due to lack of organizational commitment and leadership. Some others were found to have survived using efficient organizational form and a lean staff. An analysis of 364 small businesses revealed that 60% failed due to marketing problems, 24% failed due to management problems, and 16% failed due to finance problems (Pabst, Casas, & Chinta, 2016). Despite the struggles to survive, the growth of small businesses affects local employment opportunities positively. Yong Suk (2017) found that a 10% increase in the birth of small businesses increased metropolitan statistical area employment by 1.3%–2.2% and annual payroll by 2.4%–4.0%.

The success factors of small businesses vary from region to region. A small business in Japan grew in domestic production by expanding overseas (Shohei, 2016). At the same time in Sudan, the small businesses’ sector was considered critical for the growth of the economy of the country (Dube & Dube, 2016). Another success factor is the effort taken to prioritize and develop small businesses. For example, the financial help provided by Microfinancing improved the growth of small businesses (Dube & Dube, 2016).

The support through the policies of the local government also can affect the growth and sustainability of small businesses. For example, small business policy in Australia indicates an interest on the part of the government to use small firms as a vehicle to provide economic growth (Mazzarol & Clark, 2016). Mazzarol and Clark (2016) argued that in both Australia and New Zealand, the small business sector comprised a large portion of the total business. Over 99.7% of 2.5 million active businesses are considered small businesses in Australia and 99.4% of 502,170 businesses in New Zealand are considered small businesses (Mazzarol & Clark, 2016). Thus, small businesses in both Australia and New Zealand are a significant portion of the national economy. In China, the rise of Deng Xiaoping in 1978 led to economic freedom and liberty (Mazzarol & Clark, 2016). Two world wars and the Great Depression in 1930, led the United States to provide freedom, individualism, and impartial opportunity to small businesses (Mazzarol & Clark, 2016).

In Europe, small businesses and big companies face similar challenges such as, adoption and integration of technology (Lipton & Solomon, 2017). Small businesses have problems in obtaining finances because 80% of the financial intermediation is done through the banking system (Kraemer-Eis & Passaris, 2015). However, the Euro-area banks are holding liquid loans to small businesses, which could promote more lending to small businesses and revive small businesses (Kraemer-Eis & Passaris, 2015). Small businesses in the southeastern European region have introduced a better work environment to improve performance in the work place (Prouska, Psychogios, & Rexhepi, 2016). Despite economic performance improvement, their security and reliability concerns do exist (Lipton & Solomon, 2017). In the United States, small businesses play a vital role in the economy (Asiedu & Freeman, 2007). However, small businesses in the United States are affected by large business monopolies. Hence, small businesses may plan to use more technology to remain competitive.

The definition of small businesses varies from region to region. The European Commission defined the small business as an autonomous business with fewer than 250 employees (Mazzarol & Clark, 2016). The Australian Bureau of Statistics defined firms that employ between five and 19 as small businesses, and firms that employ between 20 and 199 as medium businesses (Mazzarol & Clark, 2016). In New Zealand, firms employing fewer than 20 employees are considered small, and firms having employees between 20 and 50 are considered medium businesses (Mazzarol & Clark, 2016). In the United States, small businesses are defined as firms having employees fewer than 250 or fewer than 1500, depending on the type of industry (Small Business Administration, 2016). For this research, small businesses are defined as firms having fewer than 250 employees.

Small Business IT Innovation

IT innovation in the area of data management is a key to competitive advantage. Data can be hosted on the premises or in the cloud. Often small businesses use the infrastructure or software hosted in the cloud, thus becoming vulnerable to all cyber-attacks against the system's confidentiality, integrity, and availability. Analysis of data could reveal advanced, persistent threats, data exfiltration attempts, and lateral movement of malware from one machine to another (Stewart, 2014). Using big data security analytics on the data hosted in the cloud is one of the most convenient ways to detect advanced, persistent threats and exfiltration attempts, since the cloud provides more storage and tools to store and process data. Threat analytics platforms, antivirus vendors, spam filters, and big data analytics engines can be used to analyze data and to protect systems from being infected (Stewart, 2014). Traditional analytic solutions are found to be less suitable for the big data space most likely because of the efficiency lacking in space and time (Alsuhibany, 2016; Zhang, Shen, Pei, & Yao, 2016). The extraction of value is crucial in big data security analytics and it requires processing large volumes of data with memory and time efficiency for small businesses.

Cloud computing powered by technological advancement in Internet bandwidth availability and virtualization has become the predominant location to store and process big data. Cloud computing also comes with a few risks for its clientele. Consolidating systems and data in one large infrastructure managed by a single vendor is one of the biggest risks of cloud computing (Rigoni & Lindstrom, 2014). A related risk is also the relinquishment of control to cloud computing vendors. In the event of a breach, many customers and companies hosted in a cloud could be affected (Rigoni & Lindstrom, 2014). The top nine threats identified in cloud computing include: data breaches, data loss, account hijacking, insecure application programming interface, denial of service, malicious insider, abuse of cloud services, insufficient due diligence, and shared technology issues (Rigoni & Lindstrom, 2014).

With the increased use of IT in small businesses, information and communication systems are no longer simply an option but a necessity. Businesses have started to rely on them in their day-to-day operations. The number of devices on the network has increased, which gives more attack surface for the intruders who are increasingly motivated and trained (Stewart, 2014). Stewart proposed techniques, tactics, and procedures to survive malicious attacks. Three areas to reduce the attack surface include: (a) fortifying basic activities such as patching, identity-based authentication, and eliminating dark space; (b) creating doubt in the adversaries' mind using moving targets, honey tokens, and misinformation; and (c) analyzing data and traffic for indicators of compromise (Stewart, 2014).

Detection of malicious threats is becoming complex for the small businesses. Traditional security measures are less effective in defending advanced, persistent threats. Firewalls are not adequate to detect all threats that enter the organization through e-mails, web traffic, and socially engineered techniques (McMahon, 2014). The perimeter security paradigm that once protected safety by using firewalls can no longer completely protect the network that consists of fragile building blocks inside the network (Lindner & Gaycken, 2014). Even encryption technology cannot protect businesses completely, since encryption can protect only certain content and cannot protect everything on the system (Lindner & Gaycken, 2014). A traditional approach also does not provide all the intelligence necessary to thwart attacks before they cause destruction. For example, the time interval between zero-day attacks and the antivirus solutions detecting them could be detrimental to the network (McMahon, 2014). Although large-scale meta-data analysis can be used to detect weak signs of compromise, the attacks are now more sophisticated, organized, focused, and malicious, which are undetectable by normal defensive strategies (Mcmahon, 2014). The analysis of big data using big data security analytics tools and techniques could provide the solution for remediating the weaknesses in the traditional defense.

Big Data

Big data were originally introduced to the world of computing by Maguoulas in 2005 to define data that could not be processed by traditional relational management databases due to their structure, complexity, and size (T. Hashem, Datta, et al., 2015; Y. Kim, Y. H. Kim, Lee, & Huh, 2015). Big data represent a huge amount of data containing data about many aspects of our lives (Strong, 2014; Trifu & Ivan, 2014). This high dimensional and nonlinear big data come from various structured, semistructured, and unstructured resources and are of different types (Qiu et al., 2016). The HACE theorem defined big data as data that are not only large but also heterogeneous, having autonomous sources, complex, and evolving (Wu, Zhu, G. Q. Wu, & Ding, 2013). Tamhane and Sayyad (2015) further affirmed that, based on the HACE theorem, the key characteristics of big data include huge and diverse data sources, distributed control, and evolving complex data. In general, big data are characterized by five big Vs: volume, velocity, variety, veracity, and value (Addo-Tenkorang & Helo, 2016).

The term *volume* refers to the huge volumes of data gathered by an organization (Kim et al., 2015). Big data contains huge volumes, and the data arrive in real time making them constantly grow (Williamson, 2014). Data are being generated around the clock, and the big data storage systems have the capability to store the huge volumes of data that both the computer and the Internet have generated (Rahman & Aldhaban, 2015). Facebook's status messages and Twitter's tweets are some of the examples of big data that come almost in real time. Credit card transactions are also produced in huge numbers. Rahman and Aldhaban asserted that there are millions of credit and debit card transactions created per minute. Big data are so huge that it is almost impossible to process those using traditional transactional databases (Caldarola, Picariello, & Castelluccia, 2015). Most traditional databases handle data by loading them into memory. However, such a strategy would not work for big data since the data are too big to load into memory (Qiu et al., 2016). Wu et al. (2013) asserted that as the volume increases, the complexity of the data also increases. For example, Google alone processes 24 petabytes (1 petabyte = $2^{10} * 2^{10} * 2^{10} * 2^{10} * 2^{10}$ bytes) of data daily, and the sheer volume of data brings challenges in learning (Qiu et al., 2016). The multiscale data from individuals continues to rise, causing a huge collection of data in terms of zettabytes ($10^{21)}$ and yottabytes (10^{24}), from various devices, such as real-time imaging, point of care devices, and wearable devices (Andreu-Perez, Poon, Merrifield, Wong, & Yang, 2015). For instance, McNeely and Hahm (2014) affirmed that in 2000, only a quarter of all stored information was digital, whereas by 2013 more than 98% of the world's stored information was stored electronically. Big data continues to grow from various sources, such as e-mail, tweets, blogs, audio and video files, and chat session logs. In addition, big data generated by machines using smart counters and sensors produces data sets of enormous sizes (Sen, Ozturk, & Vayvay, 2016).

The term *variety* refers to the different types of data formats of data. For example, data stored in relational databases are structured, whereas e-mail information has a header, which is semistructured (Rahman & Aldhaban, 2015). Free voice, text, and images are categorized as unstructured data. Big data contain not only structured data, but also unstructured data, which often has free-text, posts, messages, audio, video, and sensor data (Rassam et al., 2017). Williamson (2014) found that approximately 75% of big data contain unstructured data coming from text, voice, and video. Wu et al. (2013) also affirmed that big data could contain both heterogeneous data and diverse dimensionality of the same data.

The term *velocity* refers to the speed of data that are being generated. For example, for every second, there are more than 2 million e-mails that are being sent (H. Zhang et al., 2015). The number of e-mails sent per minute has reached 204 million (Apurva et al., 2017). The data generated from devices and sensors are coming real time to the big data storage systems. Data coming from dispersed locations, such as the geo-dispersed data can be simple as well as complex, such as the analysis of video content (H. Zhang et al., 2015). Wu et al. (2013) observed that in 2012 Flickr received 1.8 million photos daily, and the data from the square kilometer array in radio astronomy, which were 100 times more than what the conventional telescopes could render, were stored offline as real-time processing of this big data was not possible.

The term *veracity* refers to the truthfulness of the data. Big data should have traceability to ensure that the data arrived are from reliable resources (Rahman & Aldhaban, 2015). Veracity also refers to the trustworthiness of big data (Andreu-Perez et al., 2015). People are usually reluctant to use data that are not authentic or accurate. Sometimes the data are incomplete and inaccurate coming from different sources, making veracity a serious concern (Qiu et al., 2016). Veracity can be ascertained using some advanced deep learning methods that handle some level of noise in data (Qiu et al., 2016).

The term *value* refers to new ways of using data (Rahman & Aldhaban, 2015). Big data can be processed using process streaming, and the structured and non-structured data can be analyzed to generate economically useful insights and benefits (Sen et al., 2016). Drawing valuable information from big data provides deep insights that can provide competitive advantage (Qiu et al., 2016). Knowledge discovery databases and data mining algorithms are used to find the necessary information often concealed in the massive amount of data (Qiu et al., 2016).

To match the growing demands of big data, engineers have increased hardware capacity to process data, increased memory to analyze data, and invented faster processing algorithms to process big data (Wilkes, 2012). There is a variety of software tools and databases that have been developed to process, store, and analyze big data. The data storage also has become cheaper, enabling businesses to store large volumes of data on commodity servers (Rahman & Aldhaban, 2015). Traditional relational databases cannot be used to store and process the unstructured data effectively (Rahman & Aldhaban, 2015). Because of the ability to store much structured and unstructured data, big data systems are capable of storing even trivial details. Big data processing goes through multiple phases, such as data generation, data storage, and data processing

(Mehmood, Natgunanathan, Xiang, Hua, & Guo, 2016). Big data analytics has the potential to find new facts about an individual by finding correlations (Young, 2015). On a similar note, Rahman and Aldhaban (2015) asserted that organizations could increase their profitability by using these additional facts derived from big data processing.

Big Data Sources

Big data characterized by the five Vs, namely volume, variety, velocity, veracity, and value, originate from various resources such as the Internet, mobile devices, social media, geospatial devices, and sensors (T. Hashem, Datta, et al., 2015). Major sources of big data in business enterprises come from websites, tweets, and sentiment analysis from social networks, such as Google and Facebook (Caldarola et al., 2015). Other sources of big data come from less structured data, such as weblogs, web applications, mobile applications, social media, e-mails, sensors, and photographs (Wilkes, 2012; Wu et al., 2013). Social networking is another major source of big data, which has private data that need to be protected (Bardi et al., 2015). Similarly, Global Positioning Systems can generate a massive amount of data. Location-based services produce a plethora of data about an individual's location. Big data are also generated when users generate data about themselves (Musolesi, 2014). Data are generated about neighborhood events and places, as individuals travel and check-in to hotels (Musolesi, 2014). Twitter and Facebook messages are producing big data phenomenon with millions of messages and tweets (Musolesi, 2014). Third parties that mediate between businesses and customers also generate data, which are often called *innomediaries* (Caldarola et al., 2015). Mobile phones are becoming one of the greatest sources of real-time data. For example, mobile devices generate more data about employees than desktop computers and video cameras (Karim, Willford, & Behrend, 2015). Additionally, big data from mobile phones are huge and are a good source of predicting future trends (Musolesi, 2014). The enormity of big data is complex, since the data coming from the Internet, social networks, communication networks, and transportation networks are stored in big data (Wu et al., 2013). Wu et al. (2013) further argued that the value of big data is their complexity, which is represented by mixed data types, complex semantic associations in data, and relationship networks in data.

Smart homes, which are homes equipped with smart devices and sensors that can communicate through the Internet, are another source of big data. A single smart home can generate thousands of transactions daily (Bouchard & Giroux, 2015), and the big data storage and processing mechanisms can be used to process them, to find meaningful patterns in the cluster of real-time data. Additionally, big data can be used to store the activities of daily living for future analysis. Bouchard and Giroux (2015) predicted that there could be a network of smart homes in the future and that the data generated from it could be huge. It is predicted that the real-time analysis of data from smart homes could be used to assist individuals with reduced mobility or autonomy (Bouchard & Giroux, 2015).

The results of scientific experiments that produce a large amount of data are another source of big data. Research, such as the Large Hadron Collider experiments, and research using the Square Kilometre Array Telescope, the Slogan Digital Sky Survey, and the Large Synoptic Survey Telescope produce a massive amount of data (Bardi et al., 2015). The data generated from these devices are stored using big data infrastructures for big data analytics.

Big Data Analytics

Big data analytics are the process of inspecting, cleaning, and eliciting useful information using software and hardware tools (Jackson, 2014). Big data are a huge source of data for extracting information that can provide a sustainable competitive advantage for businesses (David, 2014; Tallon, Ramirez, & Short, 2013). Tallon et al. (2013) posited that information grows in value with greater use. By using the insights gained from big data analytics, businesses can increase their operating margins and achieve outstanding performances against competitors (Tan, Zhan, Ji, Ye, & Chang, 2015). Williamson (2014) asserted that big data analytics could be an enabler or a disrupter. Review of a few past studies indicated that there were significant benefits to customers who used big data (Cui, Yu, & Yan, 2016; Englmeier & Murtagh, 2017; Rahman & Aldhaban, 2015; Verma, 2017). For example, big data have been used in Singapore to foil terrorism and also have enabled the government to personalize public services to a selective population effectively (Williamson, 2014). Big data modeling has helped the Danish company, Vestas Wind Systems, to maximize power generation and also to minimize costs (Rahman & Aldhaban, 2015). Big data analytics are used in the government sector and the area of health-care, smart services, and the Internet of Things (Rahman & Aldhaban, 2015). Big data mining is being applied to marine observations to draw meaning out of the large volumes of data. For example, the marine observation satellite from NASA records movements in the ocean and the sea surface height (Huang et al., 2015). Big data can help to identify marine-related events, and forecast disastrous weather events. In some cases, data analytics without careful verification could bring defective results that are risky to the business. For example, if big data analytics in Digital Disease Detection falsely identify a location of a disease, tourism and economy of that region could be adversely affected (Vayena, Salathé, Madoff, & Brownstein, 2015).

Big data analytics in medicine. Big data analytics also have found significance for medicine by detecting life threatening problems that could save lives (Marr, 2015). Big data analytics can be used to predict the outbreak of diseases, determine the price of tickets, and predict future events based on past historical data analysis, such as election events (Bardi et al., 2015). Simple queries using Google Flu predicted flu-like sicknesses (Viceconti, Hunger, & Hose, 2015). Andreu-Perez et al. (2015) observed that by using big data analytics, different pieces of information about patients from different sources, such as genomics, proteomics, imaging, and long-term sensing, can be stratified to provide personalized services to patients. Big data can be used to address other problems in modern healthcare, by simulating using a virtual physiological human (Viceconti et al., 2015). Analytic approaches are also used in medicine, to diagnose and prevent cancer by predicting outcomes (Basole et al., 2015). Big data analytics can be used to extract clinical data to find patients that have the same pattern of symptoms. Qiu et al. (2016) affirmed that companies use deep learning techniques to leverage learning from big data for competitive advantage.

Contrarily, White and Brenkenridge (2014) noted that big data may not provide all the information that one is looking for, and big data encourage the risk of finding patterns that do not exist, often termed as the practice of *Apophenia*. Big data also runs the risk of having noise and lacking veracity. White and Breckenridge warned that big data analytics should not be the panacea of data problems but rather as another set of tools and techniques to mine data.

Big data were analyzed to determine the behavioral patterns and the emotional state of human beings (Musolesi, 2014). For example, mobile big data mining can be used to predict crime possibilities and law violations (Musolesi, 2014). In addition, data from mobile devices and from social media sites can be used to predict individual personality traits. Lambiotte and Kosinski (2014) reported that big social data have the potential to predict a five-factor model of personality, which is a set of traits including openness, conscientiousness, extroversion, agreeableness, and emotional stability. It is also possible that big data analytics can be used to extract valuable information, such as the likes and dislikes of individuals to promote targeted advertising (Lambiotte & Kosinski, 2014).

Big data analytics in engineering. Big data analytics have found their use in software engineering where high-availability infrastructures are required. For example, practitioners have faced scarce storage, limited scalability, and inadequate privacy while processing operational logs (Miranskyy, Hamou-Lhadj, Cialini, & Larsson, 2016). Big data infrastructures have enabled real-time processing of logs that can reach tens of gigabytes or even terabytes, thus eliminating the need for excessive storage and scalability (Miranskyy et al., 2016). Such accomplishment is possible because of the enabling technologies of big data, such as MapReduce, Hadoop, cloud computing, matrix recovery, cognition, ontology, and semantic (Qiu et al., 2016).

Big data analytics have been possible because of great support from the underlying network, as it implies a heavy flow of traffic between different systems. Cui et al. (2016) asserted that big data technologies also can be used in software-defined networking (SDN) to enhance security. Cui et al. agreed that there are similarities between SDN and big data, and a collaborative look at their designs can help each to perform better in aiding small businesses in network security. For example, SDN can manage the network efficiently to improve big data applications, and big data analytics can bring benefits to SDN by detecting and defeating security attacks (Cui et al., 2016). Big data are also used to enhance cloud security by powering intrusion detection systems through Hadoop infrastructure. Z. Tan et al. (2014) affirmed that the MapReduce framework provides a distributed and parallel infrastructure to implement effective and collaborative intrusion detection systems. Such powerful technologies can be utilized by small businesses if big data security analytics are implemented to detect and thwart security threats.

Big data analytics in business. The growth of big data now necessitates growth of new skills for companies to remain competitive. Rahman and Aldhaban (2015) affirmed that analyzing big data requires a new set of tools, technologies, and people with new skill sets for data visualization. Although big data processing software is distributed through open source forums, the skills needed to use the big data tools are rare and expensive. Lack of adequate skill sets is defined as one of the barriers to big data initiatives (Rahman & Aldhaban, 2015). Some organizations have learned to use a few skilled big data practitioners to develop services that encapsulate big data operations (Kim et al., 2015). Many companies are expecting employees with skills and expertise to handle large volumes of data used for predictive analytics (Earnshaw, Silva, & Excell, 2015). Small businesses could have the opportunity to hire employees with these new skills, thus bringing significant positive social change into the IT community.

Both in small and big businesses, analyzing big data by identifying patterns and correlations can lead to faster and better decisions (Adolph, 2014; Tallon et al., 2013). For example, a television firm was able to analyze unstructured big data and obtain information about the shows that are popular, and therefore the value of a commercial spot (Prescott, 2014). Big data analytics also are used to improve business activity monitoring, which provides insights into business performance (Vera-Baquero, Colomo-Palacios, & Molloy, 2016). However, challenges exist around data storage, analytics, and integration of big data. Caldarola et al. (2015) noted that big data have to be mastered to avoid collecting a huge and meaningless pile of data. Additionally, the laws have not evolved along with the pace of technology (Bardi et al., 2015). Organizational and legal policies need to be implemented to take care of protecting privacy data.

It is not having big data that makes the difference, but the processing and the insights drawn from the big data that matters (Williamson, 2014). Rubinstein (2013) posited that big data processing could provide previously unknown and possibly useful information from the massive tsunami of data. Insights are drawn from *data mining* and *data analytics,* in which advanced processing methods are used to analyze the patterns, and trends contained in the big data (Williamson, 2014). Data mining extracts interesting patterns or knowledge from big data (Hussein, Hamza, & Hefny, 2013). It is only through big data analytics, which provide a fusion of open source intelligence and social media analytics, that we can obtain intelligence about advanced, persistent threats (McMahon, 2014). Agility and big data analytics, with cloud-based infrastructures, provide a network fortress to detect enemies that are constantly moving, and maneuvering inside the corporate network (McMahon, 2014). Small businesses can use big data analytics for detecting threats that could not be handled through traditional threat prevention mechanisms.

Big data analytics in real-time processing. One of the significant features of big data is the ability to process unstructured data in real time and leverage them for decision making (Gold, 2014; Pigni, Piccoli, & Watson, 2016). Everyday data continues to increase with the prediction that there will be 4.1 terabytes of data generated per square kilometer in 2016 (Zhu et al., 2015). Real-time data come from countless devices connected to the Internet, which need to be captured, and utilized (Chen & Zhang, 2014). As all the data cannot be stored in memory, big data are processed as they arrive, in a real-time manner. For example, the Marketing department in small businesses can receive customer feedback in real time through big data analytics. Truong, Bui, and Tran (2015) found that distributed systems, such as GPSInsights, can handle the enormous volume of data in real time, and analyze them using Spark Streaming and Apache Storm, which are popular open-source frameworks for distributed processing. Big data in real time contain not just old static data, but also dynamic and continuously changing data (Musolesi, 2014). For example, real-time data from telephones produce more detailed knowledge about people, things, and events in different parts of the world (Musolesi, 2014). Real-time data coming from sensors and global positioning systems are enormous, providing a wealth of information about people and things (Musolesi, 2014).

One of the biggest features of big data analytics is the ability to process real-time logs from software applications. Big data hosting platforms provide powerful infrastructure to process tens of terabytes of operational logs. The logs are generated quickly, as with the Internet of Things, there could be millions or billions of devices pouring a vast amount of real-time data into the network (Zhu et al., 2015). Pfleeger (2014) asserted that there are 7 billion people on earth and nearly 7 billion mobile phones. Pfleeger further asserted that 6 billion e-mails are sent hourly with 1.2 petabytes of data crossing the world through the Internet. Intelligent products connected to the Internet become vehicles for sophisticated software functionality that collect and transmit data autonomously (Bello & Zeadally, 2016). Such data can bring information about attackers, botnets, and advanced, persistent threats that can be processed to detect intruders. Small businesses can use the powerful big data mining infrastructure to detect advanced, persistent threats using the logs collected over a period.

Small businesses also can benefit greatly from real-time monitoring using big data analytics. For example, big data can be used to detect a person’s future activity, based on an analysis of the past activity. Ferguson (2015) argued that modern law enforcement could combine several databases, such as law enforcement databases, third-party tools, biometric, and facial recognition software to query the records matching the person on the street, and predict a possible crime or burglary even before it happens. Large data are being analyzed to produce small data, accurate enough to prevent a possible crime or theft.

In modern healthcare, real-time imaging and continuous monitoring of individuals produce a huge amount of structured and unstructured data (Andreu-Perez et al., 2015). Big data arriving real time are used to monitor blood pressure and heart beat, while individuals run on the thread-mill using wearable digital devices. Andreu-Perez et al. (2015) observed that big data could be a valuable resource to improve health service and reduce healthcare costs, but also raises challenges regarding privacy, identification of the individual even when anonymized, ownership of data, and stewardship of data. For example, big data seem to have challenges in data-control, privacy, and quality control (Huang et al., 2015). The term *privacy issue* refers to the lack of anonymization and inability to protect personal data (Rubinstein, 2013). One of the reasons for loss of privacy is also outsourcing where data are uploaded to a cloud (Mehmood et al., 2016). The data in the cloud are likely to be accessed or lost due to data breaches. Mehmood et al. (2016) reported that multi-tenancy is another big data issue since malicious users in the same environment can illegally access data belonging to others. Big data are increasing every day and organizations must consider ways to manage their data against all privacy challenges (Sutikno, Stiawan, & Ibnu Subroto, 2014). Sutikno et al. (2014) argued that organizations must protect sensitive data by using cryptography and granular access control methods.

There are other challenges in processing real-time data. The data collected through social media are heterogeneous and may not be conducive to making decisions.

Constantiou and Kallinikos (2015) affirmed that big data contains dangerous, trivial, and messy data. The accuracy of prediction should be improved to enable decision makers to trust information derived from big data (Constantiou & Kallinikos, 2015). Real-time applications, such as navigation, social networks, biomedicine, and the Internet of Things require faster processing, which big data struggle to accommodate (Chen & Zhang, 2014). For example, location-based services are growing with the usage of smart phones. Location-based services require the user to reveal the location, and to identify points of interest around that location. Applications running on Android phones can send location information through the regular phone network, thus acting as a global positioning systems' sensor (Magtoto & Roque, 2012). Such implementations demonstrate that the real-time tracking and storing of a continuous stream of data using big data infrastructure are already in use. Despite the challenges, novel authentication mechanisms have been developed to implement privacy-preserving algorithms by using location-based services (T. Hashem, Datta, et al., 2015).

Big data analytics and privacy. Bardi et al. (2015) affirmed that novel solutions, such as rule-oriented data could be used to enhance privacy and security of big data. The use of federated grids, intelligent clouds, and distributed rules engines are a positive step toward securing big data (Bardi et al., 2015). Big data privacy issues have been addressed during different stages of the big data life cycle including data generation, storage, and processing (Mehmood et al., 2016). During data generation, data can be falsified, and true information can be hidden through anonymization techniques. During data storage, encryption can be used, such as attribute-based encryption and identity-based encryption

(Mehmood et al., 2016). Mehmood et al. (2016) reported that privacy-preserving data publishing could be used to protect data during the data processing phase.

Big data processing not only extracts information from existing data but also brings out new information from the stored big data. Big data analytics can produce personal information, especially while processing *inter* big data, which spans over multiple organizations or multiple social networks (Han & Liu, 2018). With improved mining algorithms, it is possible to find the exact footprint of an individual based on unclassified information (Bardi et al., 2015). Big data users expect privacy including data privacy, index privacy, keyword privacy, trapdoor unlinkability, and rank privacy (Chen et al., 2016). Tari (2014) observed that the key challenge in big data management is to assure the confidentiality of the privacy-sensitive data, while the data are stored and processed in the cloud.

Wu et al. (2013) observed that the common anonymization approaches, such as suppression, generalization, perturbation, and permutation could be used to generate anonymized data, eliminating personal information. Wu et al. further affirmed that once the data are anonymized, it can be freely distributed across without fear of revealing personal information. One of the privacy-preserving techniques is *privacy-preserving aggregation*, which uses homomorphic encryption when an algorithm is used to encrypt big data using public key encryption method (Wu, Yang, H. Wang, C. Wang, & R. Wang, 2016). The encrypted data are cumbersome to operate while processing big data. Wu et al. (2016) noted that operating over encrypted data is inefficient. The big data can be reduced in size by using a well-known data aggregation technology which helps in data transmission and speed. Big data can also be anonymized using quasi-identifiers with attributes shown to the public while keeping personal information confidential (Shamsi & Khojaye, 2018). Past research indicated that removing personal information through de-identification is more effective and flexible. However, it is possible that the data that were de-identified can be reidentified using correlation of data sets containing information found on social networks, blogs, and tweets (Shamsi & Khojaye, 2018). Hence, de-identification is not a complete mechanism for ensuring big data privacy.

Big data analytics for the internet of things. Since the advent of the Internet of Things, video content generated by numerous devices and high-data-rate sensors is becoming common (Satyanarayanan et al., 2015). Body-worn cameras produce video content, and transmit them to local cellular towers. Satyanarayanan et al. (2015) found a new technique to store the actual video in a cloudlet instead of the cloud. Cloudlets can send videos with their private contents encrypted, and such videos are called denatured videos (Satyanarayanan et al., 2015). Satyanarayanan et al. further reported that denatured videos contain two output files, a low frame-rate video and an encrypted video. The Cloudlet architecture allows for local storage of videos without uploading everything to the cloud, thus protecting privacy and at the same time providing value to the content analyzers.

Big data might contain sensitive information, such as personal, identifiable information (Kshetri, 2014). Rassam et al. (2017) affirmed that personal information, which reveals a user’s identity and genomic data, are collected along with consumer–related information. As many devices are connected using the Internet of Things, automatic monitoring of devices brings more exposure to private data. Earnshaw et al. (2015) argued that personal and corporate information about products and services can be stolen. The same platform that was built to increase connectivity could also be used to steal customer information. Van de Pas and Van Bussel (2015) further enunciated that the information and communication systems have ability to process big data, and the protection of citizens' privacy cannot be completely secured. Nearly 75% of the Internet users have mild to serious concerns about privacy, which have to be addressed with laws and regulations that can be commonly interpreted by the citizens in social environments (Van de Pas & Van Bussel, 2015). However, the privacy of data could be achieved through unlinkability, transparency, and intervenability (Perera, Ranjan, & Wang, 2015). Unlinkability can be achieved by limiting the data transmission to outside third parties. Transparency allows the data owners to know what is being sent, and intervenability allows for the ability to withdraw information at any time (Perera et al., 2015).

Traditional private key encryption and identity-based encryption does not protect the data effectively, since some sensitive information still can be leaked to the public while using private key encryption in the big data environment (Liang, Susilo, & Liu, 2015). There is a need for a fine-grained cipher text exchange between servers that use big data. The access control mechanisms should allow the content owner to specify the recipients easily. Liang et al. (2015) found a cipher text sharing technique that supports not only data encryption, but also supports anonymity, multiple receiver-update, and conditional sharing. Fine-grained sharing using cipher text enables big data to preserve privacy while processing a huge amount of user data. Perera et al. (2015) also reported a few strategies to overcome privacy issues in the Internet of Things' environment, such as minimize design strategy, onion routing, hide design strategy, and aggregate design strategy. Minimize design strategy recommends releasing only minimal data to third parties. Onion routing embraces anonymous communication strategy. Hide design strategy proposes hiding data. The aggregate design strategy recommends sending only aggregate data to third parties (Perera et al., 2015).

One of the emerging research topics looks at how to protect sensitive information through privacy-preserving data mining (Hussein et al., 2013; Xu, Jiang, Wang, Yuan, & Ren, 2014). Xu et al. (2015) argued that privacy information could be protected by identifying methods to protect sensitive information and by classifying the users who access data into four different types: data provider, data collector, data miner, and decision maker. By distinguishing the responsibilities of users and by empowering them with the task of hiding sensitive information, privacy information can be protected. Xu et al. further posited that during data mining, the data miner could use mining algorithms that can extract useful information without invasion of privacy. Other methods of privacy-preservation include perturbation-based solutions and cryptographic solutions. However, in perturbation-based solutions, it is possible that data can be leaked if the information is not perturbed sufficiently (Vaidya, Shafiq, Fan, Mehmood, & Lorenzi, 2014). Randomization is a technique that has been used for many data mining tasks, such as classification, regression, and ranking (Vaidya et al., 2014). Random decision trees are very useful when selecting only specific nodes that need to be hidden. Randomization of encryption and decryption also makes the task of predicting the data that are encrypted almost impossible. A combined solution using both a cryptographic solution and random decision trees has been found to be effective in preserving the privacy of big data (Vaidya et al., 2014).

Informed consent is the ability to turn the collection, handling, and processing of a customer’s data upon customer’s consent, while anonymization is the promise to maintain privacy and decouple all the personally identifiable information. Barocas and Nissenbaum (2014) noted that the problem of informed consent and anonymization are difficult to achieve. The behavior of a few people can be used to target a larger audience, and to individuals who have opted out of sharing confidential information. As in the case of the Target company, that sent out offers to a pregnant woman even before her family knew about the pregnancy, big data brings the ability to discover data from easily observable and accessible qualities while honoring informed consent (Barocas & Nissenbaum, 2014).

Government policies for retaining consumer information are of paramount importance, especially in the light of big data. Reliable public information can be useful for big data enthusiasts to draw out more hidden information through disambiguation. The data integrity principle of the European Union requires that inaccurate and incomplete information of individuals collected should be erased or rectified. The U.S. Privacy Act mandates maintaining accurate, relevant, and complete information for each individual (Waterman & Bruening, 2014). Washington (2014) affirmed that the categories of information provided by the government could be used as an authoritative source for obtaining localized data. For example, using zip codes and other peripheral information collected through big data, the tools can predict or extract personally identifiable information. The government also produces categorical information and inferential statistics and makes them available on government websites for informing the public (Washington, 2014). There is a concern that such reliable information can be used to find more personal, identifiable information.

Privacy-preservation techniques have been difficult to develop due to the volume of big data and the scalability issues related to processing big data using conventional anonymization algorithms. However, X. Zhang, Yang, Liu, and Chen (2013) developed a scalable two-phase top-down specialization approach using MapReduce tools to solve privacy-preservation issues in big data. By anonymizing private data, big data processing using cloud computing becomes useful to share the derived information. MapReduce tools coupled with cloud provide the capability for applications to mine big data and resolve the privacy issues in big data processing (Zhang et al., 2013).

However, privacy can be traded for incentives (Xu, Jiang, Chen, Ren, & Liu, 2015). The more the people are willing to sacrifice for privacy, the more they can be rewarded, and it is a compromise between privacy protection and data utility. Xu et al. (2015) reported that anonymization causes reduced usage of big data. If privacy information can be protected, it is possible to increase the adoption of big data. Hence, privacy can be treated as a type of good that can be auctioned by giving compensations to data owners (Xu et al., 2015). This approach to privacy protection allows different individuals to choose their levels of privacy protection, since some might prefer to protect more information than others. Xu et al. proposed the contract theoretic approach wherein a high level of anonymization provides data owners more privacy and less compensation, and, therefore, less data utility. In addition, distributed solutions have been developed to allow mining of big data while preserving privacy (Vaidya et al., 2014).

Big Data and Security Analytics

Organizations are faced with information security issues almost every day. Big data can be used as a resource to equip cybersecurity experts with insights about the intruders, advanced, persistent threats, and cyber criminals. Manually analyzing big data for important cybersecurity events is almost impossible (Kantarcioglu & Xi, 2016). Hence, big data security analytics are required to provide insights from the massively gathered security incident big data that include system logs, vulnerability scans, firewall logs, and file integrity monitoring tools.

When addressing information security, organizations can utilize a multitude of machine data that contain information about intruders and evidences of advanced, persistent threats. However, this can be accomplished only if the organization has big data analytics capability (Gupta & George, 2016). The capabilities are categorized into human, tangible, and intangible types (Gupta & George, 2016). Gupta and George (2016) found that the tangible capabilities include data, technology, and other basic resources, the human capabilities include managerial and technological skills, and the intangible skills include data-driven culture and intensive organizational learning. These resources must be in place to use big data security analytics and draw meaningful insights from the data. Technological storage, such as Hadoop and NO SQL are expected in big data-driven organizations, and data visualization tools, such as Tableau and SAS Visual Analytics are common (Gupta & George, 2016).

Cybersecurity incident teams can follow certain threads of security attacks and correlate security events with threat intelligence to identify the hacker as soon as possible. To effectively follow the thread, comprehensive logging is a must, which requires detailed logging of packets on all ingress and egress traffic. Network logging, including encrypted traffic, such as SSL/HTTPS/TLS, produces an enormous amount of data which cannot be easily stored or analyzed using traditional databases. Instead, the analytic tools used for big data could be used to analyze these security related logs.

Traditional security information and event management tools, although they can alert unusual activity, can store data only into relational databases that are too big and clunky to query the data. Flat files of machine data are preferred over relational data models. The evolution of big data and analytic tools to process them has made it possible for a new platform called *big data security analytics* to analyze security threats using security event logs (Cárdenas et al., 2013). Correlating security events with security logs could help to track down the security breaches including advanced, persistent threats. By using people, tools, and processes in a meaningful way, it is possible to correlate events and identify the intruder efficiently and quickly.

The use of new devices, such as tablets, smart-phones, and others, has become pervasive. Many employees are given a choice to have their device at work using the bring-your-own-device strategy (Kruidhof, 2014). Information and communication technologies have enabled employees to choose their own device and link it to the network and operate on the corporate infrastructure. Such preferences increase the network connections and increase the surface attack area. Also, organizations have learned to provide big data services to users by exposing application programming interfaces (Kim et al., 2015). Kim et al. (2015) observed that combining cloud computing with big data has enabled organizations to provide resources that are infinitely scalable for big data analysis. There are virtually unlimited resources available, to process a large amount of data using features of the cloud, such as rapid elasticity (Al-Dhuraibi, Paraiso, Djarallah, & Merle, 2018).

Small Businesses and Big Data Security Analytics

Big data have been found to be crucial to the competitive edge of businesses. Caldarola et al. (2015) posited that many managers from the public sector considered big data a strategic tool to make better decisions regarding spending and providing service to the public. The advantages of using big data analytics are increasing. For example, the U.S. Department of Agriculture reduced fraud by 60% by using big data analytics (Caldarola et al., 2015). Using the intelligence provided by big data analytics, Starbucks expanded their business to hundreds of stores (Rahman & Aldhaban, 2015). However, Miao and Zhang (2014) argued that while big data bring relative advantage, such as bringing new data view, changing tools and methods, and more social change, big data also tend to have issues concerning security. Using DOI theory, the adoption of big data security analytics can be predicted by evaluating the perceived attributes of innovation including relative advantage, compatibility, complexity, observability, and trialability (Rogers, 2003).

Relative advantage is one of the perceived attributes of innovation that needs to be adopted. Big data analytics have gained acceptance as a provider of great benefits and relative advantage. The ability to analyze unstructured data in addition to traditional relational data usually provides more value to the organization. However, the current analytics infrastructure has not been very helpful, and hence better analytics infrastructures based on a deduction graph have been proposed to gain a competitive advantage in areas, such as supply chain management (Tan et al., 2015). Many organizations should also know what information they need to create more value to gain advantage (K. H. Tan et al., 2015). Big data have high operational and strategic potential in generating business value. However, there seems to be no empirical research to prove the value or relative advantage of big data analytics (Wamba, Akter, Edwards, Chopin, & Gnanzou, 2015).

Another perceived attribute of innovation is the *compatibility* of the innovation with the existing innovations. Big data security analytics need to be compatible with other technologies to provide easy adoption. Big data uses current technologies and have become an enabler of improved decision making for enhanced firm performance (Wamba et al., 2015). For example, cloud computing infrastructures provide the basic infrastructure for analyzing large distributed files, which are leveraged by big data technologies (Kim et al., 2015; Qiu et al., 2016). More empirical research is required to validate the compatibility of big data technologies.

Another perceived attribute of innovation is the *complexity* of innovation. There is an inherent complexity in processing big data, which contain both structured and unstructured data (Apurva et al., 2017). It is hard to extract and manage unstructured data since variety brings more complexity (Gil & Song, 2016). The World Wide Web provides millions of data tables with structured data (Gil & Song, 2016). Social networking, tweets, and sentiment analysis provide heterogeneous and complex data (Caldarola et al., 2015). Big data processing is thus complex in data capture, storage, analysis, and visualization (Gil & Song, 2016).

Other attributes of innovation include *observability* and *trialability*. Both of these attributes are manifested in big data implementation because of the ubiquitous presence of cloud computing and hosting of big data tools and packages, which are easily downloadable and used, by both novice and experienced professionals. Cloud computing environments, such as Amazon web services provide servers, storage, and computation environments to execute big data applications in cloud environments (Feller, Ramakrishnan, & Morin, 2015). Thus, the observability and trialability of big data tools and services have been facilitated by Amazon web services and other cloud computing environments. There seems to be no recorded study of the adoption of these complex big data technologies although the benefits are easily observable and trialable.

Verma (2017) found that the adoption of big data services is slow, especially in manufacturing firms in India. Although Verma analyzed the adoption of big data services among manufacturing firms in India, and Powelson (2012) analyzed the adoption of cloud computing in Arizona, there are few studies related to the adoption of big data security analytics, especially concerning small businesses in the United States. Despite the fact that the relative advantage of big data security analytics for small business is significant, specific information about the adoption of big data security analytics by small businesses in the United States seems to be lacking.

Summary and Conclusions

Advanced, persistent threats can be recognized only by analyzing logs over an extended period. As botnets, denial of service, phishing, malware, and website threats continue to attack corporate networks, and security is becoming increasingly important to both small businesses and big organizations (Gupta et al., 2017). The voluminous logs that are generated each day can be processed only through the use of big data analytic tools. Also, the big data analytic tools can be used to analyze security threats and alert the security professionals in the company. As many small businesses are contracted by big businesses, the attackers use small businesses' networks as springboards to get into the big businesses' networks. In protecting the networks of both small and big businesses, it is important to secure organizational assets as well as the virtual private networks that connect small businesses to their big client organizations. The perceived innovation diffusion attributes by small businesses can be used to predict the adoption of big data security analytics, which could help to detect and prevent advanced, persistent threats. The literature review indicated a scarcity of research related to the adoption of big data security analytics among small businesses. Hence, this research could improve the adoption of big data security analytics by studying the relationship between the perceived attributes of innovation diffusion and their adoption. In Chapter 3, I included the research design, population, sampling methods, procedures for recruitment, instrumentation and operationalization of constructs, and the data collection process to be used.

Chapter 3: Research Method

Introduction

The purpose of this quantitative correlational study was to examine ways to increase the adoption of big data security analytics among small businesses in the United States by examining the relationship between small business leaders' perceptions of big data security analytics attributes and their adoption. The increase in adoption could detect and prevent advanced, persistent threats from malicious resources and improve the confidentiality, integrity, and availability of data among small businesses. In subsequent sections in this chapter, I will cover the research design and the rationale for the selection of quantitative methodology. The use of a survey instrument necessitating a sampling of the population and the procedures for sampling for the survey will be discussed. The procedures for recruitment, participation, and data collection will also be enumerated. Finally, I will discuss the rationale for a web-based instrument and the operationalization of constructs to provide the constructs used to prove internal consistency and external validity.

Research Design and Rationale

In this study, I measured the adoption of big data security analytics by small businesses by examining the relationship between small business leaders' perceptions of big data security analytics and their adoption of big data security analytics. The perceived attributes of innovation (i.e., compatibility, complexity, observability, relative advantage, and trialability) correlate with the adoption of innovation, such as the big data security analytics (see Rogers, 2003). Therefore, I used these perceived attributes of innovation as the variables for this study. The criterion variable was the adoption of big data security analytics. The characteristics of this project were more consistent with a quantitative paradigm instead of a mixed method approach or a qualitative approach. The scientific approach called correlational research design, which is primarily deductive, seemed to be best suited for this study. Correlational research designs can be very robust and can be replicated for subsequent studies when the sample is large enough and the measurement is reliable (Schoonenboom, 2017).

Experimental designs are suitable for situations where manipulation or intervention is needed to observe the study, whereas the nonexperimental designs are suitable where humans can be observed spontaneously and knowledge can be gained just through observation rather than an experiment (Hansson, 2016). The nonexperimental design provides an alternate approach to examine the opinions of a selected sample such that the results may be generalized to a population from the sample studied (see Nachmias & Nachmias, 2008). Therefore, I chose a quantitative study with a nonexperimental design over an experimental design for this study. Also, because the variables representing perceived attributes of innovation can be measured using numerical variables and are not causal-comparative, a correlational research design was used to determine the degree of effect of one or more variables on the adoption and to compare and analyze relationships between the variables in this study. Moreover, there was only one group of the population that was surveyed; hence, the contrasted-group designs (see Nachmias & Nachmias, 2008) were not applicable to this study.

Statistics help to analyze the relationship between perceived attributes of big data security analytics and the adoption of big data security analytics offering sophisticated statistical insights (see Spielgelhalter, 2014). Using the theoretical lens of Valier et al. (2008) and Powelson (2012) related to the adoption of IT innovation during the prediffusion stage, I examined the relationship between perceived attributes of innovation and the adoption of innovation. The preferred correlational design was consistent with the designs used in past studies (see Powelson, 2012). I completed the quantitative analyses with appropriate statistical tests using the Statistical Package for the Social Science (SPSS) considering the normalization of the data sets (see Field, 2013). Cronbach's coefficient alpha was computed to ascertain the instrument's validity and reliability (see Field, 2013). Descriptive statistics, such as the measures of central tendencies, dispersions, and frequency distributions, were performed on the observed data. In addition to descriptive statistics, I conducted hypothesis testing, correlation analysis, bivariate regression, and multivariate regression between the perceived attributes of innovation and the adoption of innovation. Quantitative studies are well suited for the study of relationships between two or more variables by analyzing their correlation coefficients (Azucar, Marengo, & Settanni, 2018; Sapoetra, 2017). Yilmaz (2013) posited that in a quantitative study, the researcher employs objective epistemology to observe and report the facts with detachment and impartiality.

Qualitative methods are best suited to embrace social constructivism, interpretivism, advocacy, and participatory philosophical perspectives (Powelson, 2012). Qualitative research procedures can handle formative and narrative inquiries, on-site observations, and personal one-on-one interviews (McCurdy & Ross, 2018). Also, qualitative research methods are based on constructivist epistemology and are exploratory in nature, deriving answers from open-ended, exploratory questions (Yilmaz, 2013). The intensive exploratory methods of qualitative studies are not suitable for the predictive and deterministic nature of inquiries that should instead call for quantitative methods (Powelson, 2012).

Mixed methods or qualitative methods are more appropriate when the methodology is inductive, mostly involving the search for patterns among the emergent themes in the research (Yilmaz, 2013). Mixed method research is a combination of the strengths of both quantitative and qualitative studies (Abowitz & Toole, 2010). Mixed methods are suitable to compare the results of both quantitative and qualitative outcomes, assuming there is sufficient time to conduct both quantitative- and qualitative-based inquiries. Moreover, mixed method research has gained popularity in social studies because mixed methods, although expensive regarding time, money, and energy, improve the validity and reliability of the resulting data (Abowitz & Toole, 2010). Although a mixed method approach could have also been used, the objective of this study warranted the quantitative methodology alone.

By using a quantitative cross-sectional survey design instead of a longitudinal survey design, I maximized the benefit of using a sample to predict the outcome for a large population in this study. The quantitative cross-sectional survey design included these elements: hypotheses, variables, population, sampling criteria, data collection using surveys, and statistical data analysis. A self-administered quantitative survey design is more cost-effective than structured interview strategies. However, self-administered web surveys get a lower response rate due to the user’s perception of surveys as spam e-mails (Sănchez-Fernăndez, Muńoz-Leiva, & Montoro-Rīos, 2012). I mitigated this problem by using personalized invitations to a web-based survey, which increased the retention rate (see Sănchez-Fernăndez et al., 2012). Recent studies have also shown that the response rate in web-based surveys was not dependent on a single variable but on a combination of different variables (Trespalacios & Perkins, 2016). Finally, a cross-sectional survey design strategy involves data collection at a single point in time, whereas a longitudinal survey design strategy involves data collection over an extended period (Nachmias & Nachmias, 2008). A cross-sectional survey design was appropriate for this research study because it helped to collect the perceived attributes of big data security analytics at a specific point in time, such as the prediffusion stage of big data security analytics.

Methodology

I used a positivist approach for this research, and epistemologically this was done with knowable degrees of certainty by using objectively-correct scientific methods to describe the adoption of the innovation with certainty (see Molina Azorín & Cameron, 2010). In quantitative studies, statistical inference is a method of predicting for a large population based on the results obtained on a subset of the population (Makar, 2013). In this section, I will discuss the rationale for the population and the sampling process along with the eligibility of the participants. In the final section, I will present the procedures for recruitment, participation, and data collection and describe how the instrumentation and operationalization of constructs were implemented.

Population

Population refers to the collection of units or people to which researchers want to generalize the findings (Field, 2013). I determined the theoretical population for this research through the theoretical lens of the purpose statement, and the sample selected was accessible, representing the population to which this study could be generalized. For this research, the decision makers of the big data security analytics in small businesses were considered as the target population. The sample consisted of IT professionals, who had decision-making capability regarding big data security analytics in their respective organizations. The sampling frame is the source through which a researcher can obtain access to the sample (Powelson, 2012). I considered IT professionals working for the small businesses in the United States as the potential sampling frames for this research.

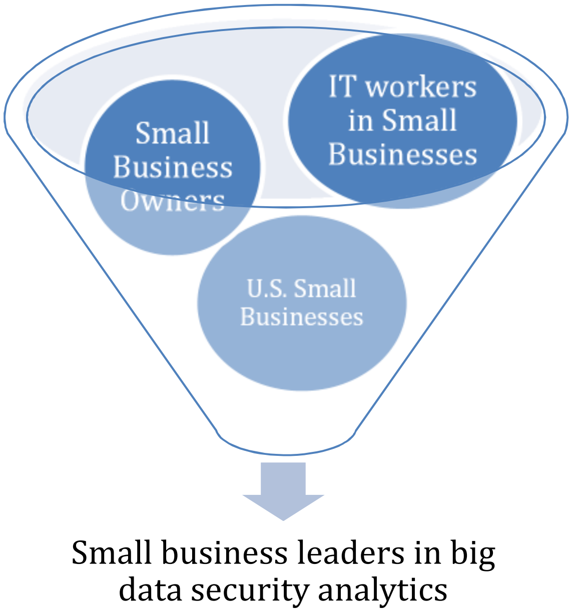

IT workers in Small Businesses
Small Business Owners
U.S. Small Businesses
Small business leaders in big data security analytics

Figure 2. Selection of sample. This figure illustrates the conceptualization of the population sampling to distill the small business leaders interested in adopting big data security analytics in the United States.

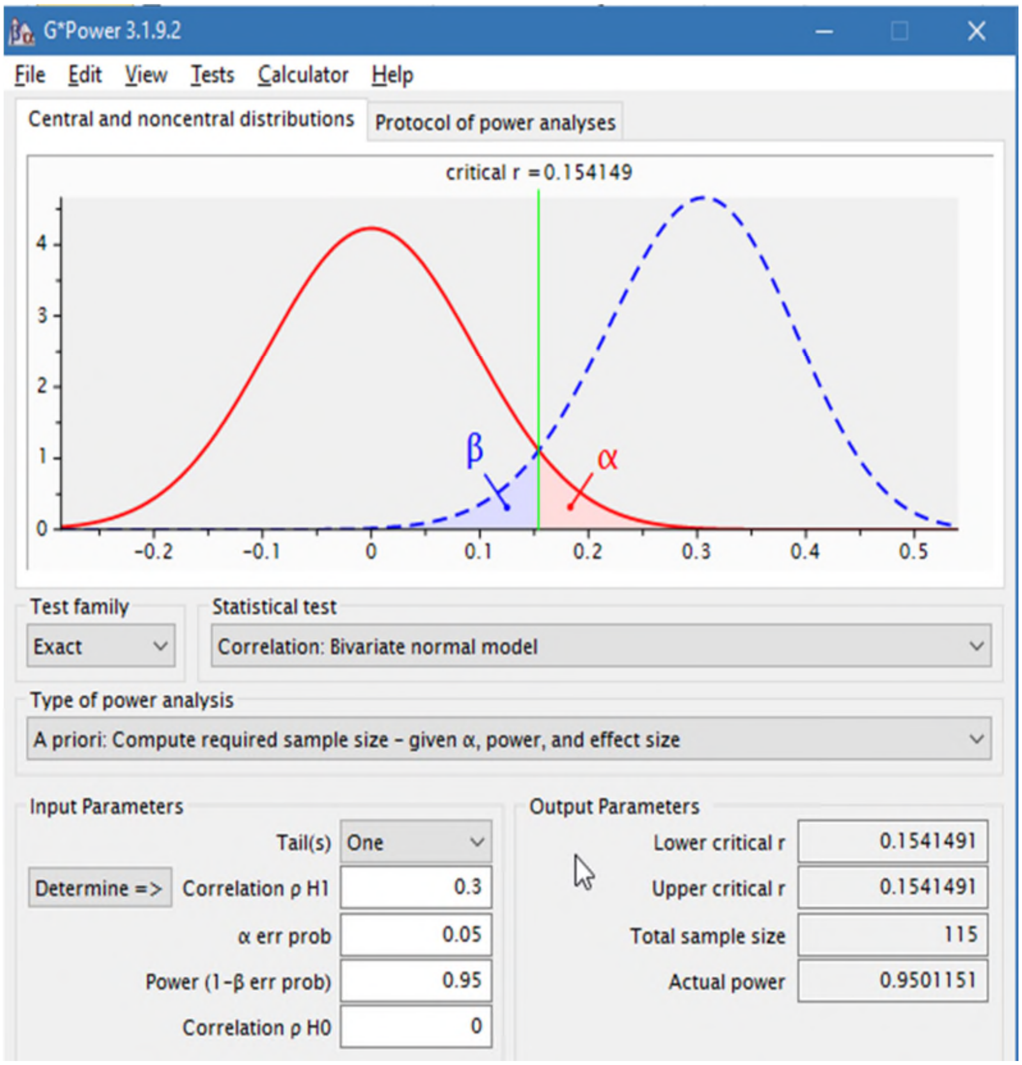
For this study, I defined small businesses as firms having fewer than 250 employees. The U.S. Small Business Administration (2016) reported that small businesses, whose size ranges from 250 to 1500, numbered around 30.2 million, comprising 99.9% of all U.S. businesses. Hence, the small businesses seem to be a vital part of the economy of the United States. Small businesses employed 58.9 million employees in 2018, totaling 47.5% of the labor force (Small Business Administration, 2018).

Sampling and Sampling Procedures

Sampling refers to the method of selecting cases from a selected population (Uprichard, 2013). The knowledge obtained from the sample may then be generalized to the population (Uprichard, 2013). I collected the samples for this study through an online survey distributed to Survey Monkey’s voluntary web participant pool. Hence, this could be considered as a convenience sampling with self-selection method, where the participants were given the option to participate in the survey. Survey Monkey’s web-based survey provided ease of access, reduced cost, the ability to answer online quickly, and the ability to access remote groups and individuals (see Wright, 2017). Although it is impractical to survey the entire population of small businesses in the United States, a portion of the population were surveyed to draw inferences about the entire population using inferential statistics (see Makar, 2013). Nonprobability sampling provides convenience but is also subject to judgment (Powelson, 2012). While the purpose of probability sampling is to extend the findings to the population, the purpose of nonprobability sampling is to know more about the cases themselves (Uprichard, 2013).

In addition, the probability sampling method requires a sampling frame which represents a large population (Uprichard, 2013). Because online surveys are more convenient in terms of cost and usage, I used web-based surveys to gather data from IT professionals who were familiar with big data, and at the same time worked for small businesses. The online survey using the voluntary web participant pool was considered a nonprobability sampling or convenience sampling because it involved self-selection and there was less chance of introducing randomness.

The sample size is the smaller collection of units representing a larger population to find the facts about that population (Field, 2013). Power analysis was used to determine the sample size. Power analysis provided also a strategy to avoid the null hypotheses when it is false. Moreover, the outcome of research was accompanied by a confidence interval to eliminate bias. The confidence interval forms a range such that obtained values from the sample that fall within that range are likely to occur within the population. Additionally, confidence intervals have been used to report scientific findings. A wide confidence interval indicates a long range of possible values while a narrower confidence interval provides a more precise estimated value (Liu et al., 2014, 2014). In the social sciences, generally a confidence interval of 95% is selected and the significance level represented by alpha is set to 0.05. The survey was hosted online and kept open until 115 or more samples were received, as suggested by the G*power software. The proposed medium effect size was .30 with a significance level of .05. The power was .95 as computed by the G*Power version 3.1.9.2 software (Faul, Erdfelder, Buchner, & Lang, 2009). The G*Power program is free, easy to use, and it provides an exact method to calculate sample size. The G*Power program is available for download from the Internet address http://www.gpower.hhu.de/en.html.



Test family: Exact

Statistical test: Correlation: Bivariate normal model

Type of power analysis: A priori: Compute required sample size – given α, power, and effect size

Input Parameters

Tail(s)	One
Correlation ρ H1	0.3
α err prob	0.05
Power (1-β err prob)	0.95
Correlation ρ H0	0

Output Parameters

Lower critical r	0.1541491
Upper critical r	0.1541491
Total sample size	115
Actual power	0.9501151

Figure 3. G*power analysis. This figure shows the required sample size.

Due to the exponential growth of internet usage, many surveys are created and administered using the online survey service provider called Survey Monkey. Web-based surveys are considered to have more benefits than traditional survey methods (Sănchez-Fernăndez et al., 2012). As responses received from web-based surveys are faster than the responses from traditional paper-based surveys, web-based surveys are preferred over traditional paper-based surveys. The web-based survey was sent to leaders of small businesses in the United States through e-mail by Survey Monkey. Instead of inviting the entire population of 30.2 million small businesses' leaders in the United States (Small Business Administration, 2018), a selected subset of the entire population who were part of the web participant pool hosted by Survey Monkey, was e-mailed and solicited for participation in the survey. Hence, a convenience sample method was used for this study, and the sampling frame was the list of participants from small businesses who are part of the voluntary web participant pool. A sufficient sample from convenience sampling was considered better than the insufficient count of responses from a large random sampling. The individuals targeted are expected to be capable of decision making, especially in the area of big data security analytics among small businesses. The individuals could be senior engineers, managers, architects, and directors or chief-level executives who are decision makers at different levels in the organization.

Procedures for Recruitment, Participation, and Data Collection

Only IT decision makers in small businesses familiar with technology were qualified to participate in the survey. The data collectors available in Survey Monkey were used to select the participants from the Survey Monkey's web participant pool, using the filters that I configured on the Survey administration page. The filters applied included:

- Decision making authorities who are part of IT Software;
- Company with employees fewer than 250;

- IT decision makers whose position included Chief Technology Officer, Chief Information Officer, Management Information System manager, network designer, network manager, project manager, business analyst, and security administrator/analyst;
- Gender including both male and female;
- Age ranging from 18 to 60+ years old;
- Area limited to all regions within the United States; and
- Screening question to qualify participants having big data familiarity.

The participants were given a web-based survey using the survey administration tool hosted by Survey Monkey. Hence, the small businesses in the web participant pool were considered the sampling frame of this study. Survey was kept open until the number of participants who completed the survey reached the desired sample size for this study.

Participants also were given informed consent using the web-based survey. Participants had direct access to the survey hosted online with the flexibility of participating in the study from their office, home, or while traveling if the Internet connection is available. E-mail was the primary mode of communication with the participant. The advantages of e-mails include speed, low cost, and elimination of geographical and social distance (Berghel, 1997). The rejection of e-mails can be reduced by using authentic service providers or organizations specialized in rendering online surveys. The estimated response rate was around 115 surveys, to achieve a medium effect size of 0.30, alpha being .05, and a power of .95 given by the G*Power program (Faul et al., 2009).

After the survey was complete, data were exported for analysis and also stored securely in Google cloud storage. Survey data were archived securely using encryption methods for 5 years to provide data safety and integrity. Data will be deleted after the 5 years to avoid any attempt to plagiarize or manipulate data.

Instrumentation and Operationalization of Constructs

Valier et al. (2008) indicated that the DOI theory could be adapted for diffusion of new technologies. Powelson (2012) adopted the instrument used by Valier et al. to customize it for the study of diffusion of cloud computing. In this quantitative study, the same instrument that was used for the study of the adoption of cloud computing by Powelson, was adapted to study the diffusion of big data security analytics by changing the reference of cloud computing to big data security analytics throughout the PreDOI survey instrument leaving the remainder of the instrument without any modification. The permission to use this published survey instrument had been obtained and is shown in Appendix C. The DOI promotes the study of adoption of an innovation and it can be used to assess the five variables of innovation: perceived attributes of innovation, type of innovative decision, communication channels, nature of social system, and change agent promotional effectiveness (Rogers, 2003). Although some small businesses might not be aware of the specific technologies related to big data, most of the decision makers should be aware of big data technology and their application to security analytics.

The constructs measured are the perceived attributes of innovation: relative advantage, compatibility, complexity, observability, and trialability, which were also used in previous research conducted by Valier et al. (2008). Powelson (2012) adapted the research by Valier et al. to study the diffusion of cloud computing by changing the reference of open source to cloud computing. Powelson’s survey instrument was adapted to complete this study by replacing the reference of cloud computing with big data security analytics without altering the remainder of the instrument. A 7-point Likert-type scale with ordinal values ranging from 1 to 7, with 1= *strongly disagree*, 4 = *neutral*, and 7 = *strongly agree* was used as the measuring units of scale for measuring the DOI. In the following section, all the variables are enumerated.

Compatibility: Perceived attribute of innovation (X_1). Valier et al. (2008) operationalized compatibility to measure the congruence between one’s experiences, values, and needs, and the propensity to adopt an innovation. Powelson (2012) continued to use this compatibility measurement to measure the diffusion of cloud computing. Compatibility of big data security analytics was measured using the Items 1 through 4 as described in the sample PreDOI survey instrument shown in Appendix A.

Complexity: Perceived attribute of innovation (X_2). Valier et al. (2008) operationalized complexity to measure the degree of difficulty in understanding an innovation. Powelson (2012) further operationalized complexity to measure the degree of difficulty in adopting cloud computing. Complexity of big data security analytics was measured using the Items 5 through 10 as described in the sample PreDOI survey instrument shown in Appendix A.

Adoption: Criterion variable (Y). Valier et al. (2008) operationalized the intent to use to measure the subject’s inclination to adopt an innovation. Powelson (2012) further used this variable to measure the subject’s adoption of cloud computing technology. Adoption of big data security analytics was measured using the Items 36 through 39 as described in the sample PreDOI survey instrument shown in Appendix A.

Observability: Perceived attribute of innovation (X_3). Valier et al. (2008) operationalized observability to assess the degree of visibility of an innovation attribute to potential adopters. Powelson (2012) further operationalized observability to measure the diffusion of cloud computing technology. Observability of big data security analytics was measured using Items 11 through 14 as described in the sample PreDOI survey instrument shown in Appendix A.

Relative Advantage: Perceived attribute of innovation (X_4). Valier et al. (2008) operationalized relative advantage to measure the advantage an innovation brings to the potential adopter of the innovation. Powelson (2012) further operationalized relative advantage to measure the adoption of cloud computing. Relative advantage of big data security analytics was measured using the Items 15 through 22 as described in the sample PreDOI survey instrument shown in Appendix A.

Trialability: Perceived attribute of innovation (X_5). Valier et al. (2008) operationalized trialability that helps to measure the ability to use an innovation. Powelson (2012) further operationalized trialability to study the adoption of cloud computing by small businesses in Arizona. Trialability of big data security analytics among small businesses was measured using Items 27 through 31 as described in the sample PreDOI survey instrument shown in Appendix A.

Powelson (2012) permitted, as shown in Appendix B and Appendix C, to reuse the survey instrument used in the study of the adoption of cloud computing. Powelson had adapted the instrument from Valier et al.’s (2008) PreDOI survey instrument that measured the diffusion of open source software. The PreDOI survey instrument from Powelson remained unaltered, except for cloud computing being replaced by big data security analytics. The dissemination of pervasive big data security analytics is very similar to the dissemination of cloud computing among small businesses and hence the instrument was well suited for the study of adoption of big data security analytics.

The survey instrument was hosted online using a web-based survey platform called Survey Monkey. The participants were invited by e-mail and text, and were asked to complete the survey as described in Appendix A. The PreDOI survey instrument contains two sections: (a) general participant items and (b) big data security analytics research items. The general information was collected from nine response items using nominal, ordinal, and interval measurements categorized as participant demographics, systems, and communications. The demographics were captured using Items 1 through 2, system information using Items 3 through 5, and communications using Items 6 through 9.

The big data security analytics survey instrument provided the participants with questions that capture the five perceived attributes of innovation along with the variable depicting adoption. The PreDOI survey instrument had 39 questions based on a 7-point Likert-type scale of ordinal values with each item ranging from 1, meaning *strongly disagree*, to 7, meaning *strongly agree*. The hypotheses were written in such a way that the greater the score for the scale items, such as compatibility, observability, trialability, and relative advantage, the more inclined is the participant to adopt big data security analytics. The hypothesis for complexity was worded in reverse to suggest that the lower the score for the item complexity, the more inclined the participant to adopt big data security analytics.

PreDOI instrument integrity. Powelson's (2012) instrument was adapted to study the adoption of big data security analytics. The original instrument was noted as reliable and valid based on the author's analysis and study. Valier et al. (2008) used the PreDOI survey instrument to study the adoption of open source software. Valier et al. validated the instrument for internal consistency and external validity by analyzing their results against previous similar studies. Powelson adapted the PreDOI instrument from Valier at al. and checked the results for internal consistency and external validity, using Cronbach's coefficient alpha and factor analysis. Modification of the instrument was limited to replacing the reference of cloud computing with big data security analytics without altering the remainder of the instrument. Hence the instrument was considered reliable and valid to study the diffusion of big data security analytics.

Similarly, the PreDOI survey instrument's integrity, internal consistency, and external validity were measured using SPSS statistic functions, such as the Cronbach's coefficient alpha and Spearman's Rho. The Cronbach's alpha 0.60 is considered to be poor, 0.70 is considered to be acceptable, and a score over .80 is good. During the study of a green fertilizer technology adoption, the PreDOI survey instrument was found to be reliable since the Cronbach's coefficient alpha ranged from 0.70 to 0.90 (Mannan, Nordin, & Rafik-Galea, 2017). The next section about data analysis provides detailed measures taken to ensure the survey instrument's internal consistency and external validity.

Data Analysis Plan

The purpose of this quantitative correlational study was to examine ways to increase the adoption of big data security analytics among small businesses in the United States by examining the relationship between small business leaders' perceptions of big data security analytics attributes and their adoption. The research question addressing this relationship was defined as follows: To what extent can DOI theory be used to encourage the adoption of big data security analytics to detect and prevent advanced, persistent threats from malicious sources among small businesses using perceived attributes of innovation including relative advantage, compatibility, complexity, observability, and trialability?

The relationship between each of the variables comprising the perceived attributes of big data security analytics and the adoption of big data security analytics was addressed by the following research questions:

Research Question 1: To what extent does the perceived attribute of innovation called relative advantage relate to the slow adoption of big data security analytics among small businesses to detect and prevent advanced, persistent threats from malicious sources?

The first hypothesis (*H*1) was postulated such that, during the prediffusion stage, the higher the level small business leaders perceive the relative advantage of big data security analytics, the greater their adoption of big data security analytics.

$H_0$1: There is no correlation between relative advantage and the adoption of big data security analytics.

H_a1: There is a positive correlation between relative advantage and the adoption of big data security analytics.

Research Question 2: To what extent does the perceived attribute of innovation called compatibility relate to the slow adoption of big data security analytics among small businesses to detect and prevent advanced, persistent threats from malicious sources?

The second hypothesis (*H*2) was postulated such that, during the prediffusion stage, the higher the level small business leaders perceive the compatibility of big data security analytics, the greater their adoption of big data security analytics.

$H_0$2: There is no correlation between compatibility and the adoption of big data security analytics.

H_a2: There is a positive correlation between compatibility and the adoption of big data security analytics.

Research Question 3: To what extent does the perceived attribute of innovation called complexity relate to the slow adoption of big data security analytics among small businesses to detect and prevent advanced, persistent threats from malicious sources?

The third hypothesis (*H*3) was postulated such that, during the prediffusion stage, the lower the level small business leaders perceive the complexity of big data security analytics, the greater their adoption of big data security analytics.

$H_0$3: There is no correlation between complexity and the adoption of big data security analytics.

H_a3: There is a negative correlation between complexity and the adoption of big data security analytics.

Research Question 4: To what extent does the perceived attribute of innovation called observability relate to the slow adoption of big data security analytics among small businesses to detect and prevent advanced, persistent threats from malicious sources?

The fourth hypothesis (*H*4) was postulated such that, during the prediffusion stage, the higher the level small business leaders perceive the observability of big data security analytics, the greater their adoption of big data security analytics.

H_04: There is no correlation between observability and the adoption of big data security analytics.

H_a4: There is a positive correlation between observability and the adoption of big data security analytics.

Research Question 5: To what extent does the perceived attribute of innovation called trialability relate to the slow adoption of big data security analytics among small businesses to detect and prevent advanced, persistent threats from malicious sources?

The fifth hypothesis (*H*5) was postulated such that, during the prediffusion stage, the higher the level small business leaders perceive the trialability of big data security analytics, the greater their adoption of big data security analytics.

H_05: There is no correlation between trialability and the adoption of big data security analytics.

H_a5: There is a positive correlation between trialability and the adoption of big data security analytics.

A correlation design was suitable for this study since the purpose of the study was to determine the extent to which the perceived attributes of innovation relate to the adoption of big data security analytics. The literature supported this design, method, and the instrument. For example, Powelson (2012) used this DOI survey instrument to study the correlation between the perceived attributes of innovation and the adoption of cloud computing. Rogers (2016) used the perceived attributes of innovation and the technology acceptance model to study the adoption of Computerized Accounting Systems among small businesses. A web-based survey instrument was suitable to study the perceived attributes of the innovation and their correlation with the adoption of big data security analytics. The use of quantitative method with correlation design and convenience sampling was appropriate for this research. The raw data collected during this research was stored and made available for a period of 5 years after publication. Data will be deleted after the 5 years to avoid any attempt to plagiarize or misuse data.

In this research, participants from small businesses were directed to surveymonkey.com to complete the survey online. After the data collection was complete, data were examined for completeness. If some surveys were incomplete, they were discarded and only the completed surveys were kept digitally for at least 5 years. Data collected through surveymonkey.com was imported into Windows-based SPSS Version 25.0 and stored in an SPSS native file system for the rest of the data analysis period. A standard 7-point Likert-type scale was used where 1 = *strongly disagree*, 2 = *disagree*, 3 = *disagree slightly*, 4 = *neither disagree nor agree*, 5 = *agree slightly*, 6 = *agree*, and 7 = *strongly agree*, to examine the relationship between the perceived attributes of innovation and the adoption. Responses to questions that were worded in reverse were coded in reverse to have consistency in interpretation. Scale items that require reverse coding include 5, 6, 7, 13, 26, 32, 37, 38, and 39. A higher score obtained will indicate a higher degree of intention to adopt big data security analytics. The surveyed data were subject to both descriptive and inferential statistics using SPSS, to study the correlation between the perceived attributes of big data security analytics and the intent to adopt big data security analytics among small businesses.

The data were analyzed for extremes or outliers. By visually screening box plot diagrams, extremes or outliers were identified and eliminated to obtain a normal distribution. Data sets that are not normal could suggest distorted relationships and Type 1 significance errors. Nonnormal data were identified using data plots, skew, and kurtosis, and outliers were eliminated to make the distribution normal.

Descriptive Statistics

Descriptive statistics was used to present demographics and general information about the population using measures of central tendency and dispersion. Descriptive statistics included frequencies, percentages, means, medians, mode, and standard deviation (Nachmias & Nachmias, 2008). The frequency diagrams were presented using SPSS. Based on the data which is imported into SPSS, both frequency measures and percentages were calculated using statistical procedures in SPSS and they were presented in a tabular format.

The mean or average is a measure of central tendency, which provides the numerical average of the observations (Nachmias & Nachmias, 2008). The mode is the measure of central tendency, defined as the most frequently occurring observation category in the data (Nachmias & Nachmias, 2008). The median is another measure of central tendency, defined as the point above and below which 50% of the observations fall (Nachmias & Nachmias, 2008). In this research, the measures of central tendency, such as the mean, mode, and median were observed using the SPSS software to measure the perceived attributes of big data security analytics.

Inferential Statistics

Inferential statistics help to make decisions or inferences about characteristics of a population based on observations obtained from a sample of the population (Nachmias & Nachmias, 2008). In this research, a sample of IT decision makers from the small businesses in the United States were surveyed using a self-administered online survey instrument. After the observations are collected, inferences were made for the entire population of small businesses using inferential statistical methods.

Multiple regression analysis is useful when there are two or more variables and the objective is to find a relationship or correlation between the variables (Rogers, 2016). The Pearson correlation was used to find out the relationship between interval variables (Nachmias & Nachmias, 2008). Bivariate linear regression was to find linearity between two variables. A multiple regression analysis model was selected to test the hypotheses. The hypotheses examined the correlation between perceived attributes of innovation and the tendency to adopt big data security analytics from a quantitative perspective.

This quantitative correlational design study included multiple linear regression with a 95% confidence interval ($\alpha = .05$). The probability standard or alpha value of .05 is an acceptable standard in academic research (Rogers, 2016) and it is also called the probability or p-value. These parameters are chosen to allow for only 5% chance of a Type I error occurring. Type 1 error refers to the incorrect rejection of a true null hypothesis, and a Type II error refers to the failure of rejecting a false null hypothesis (Rogers, 2016). If the statistical analysis shows a $p \leq .05$, the null hypothesis will be rejected. The strength of association between the variables can be found also using Cramer's V with a lower limit of 0 and an upper limit of 1. The higher the value in the range from 0 to 1, the stronger is the association between the selected variables. Sample size was calculated using G*Power 3.1.9.2 with the generally accepted power of .95. The effect sizes of *small* 0.1, *medium* 0.3, and *large* 0.5 were used. Based on prior studies related to the adoption of a technology, a medium effect size of 0.3 is proposed for this study (Powelson, 2012).

Multiple regression analysis is valid when there is linearity between the selected variables. The scatter plots produced through SPSS were used to validate linearity. Cronbach's alpha, based on classical test theory, is an intraclass correlation coefficient frequently used to measure internal consistency (De Vet, Mokkink, Mosmuller, & Terwee, 2017). The internal consistency of the survey data was determined by Cronbach's alpha values obtained through statistical analysis by SPSS. An alpha value of .70 is considered satisfactory (Rogers, 2016). Multinomial regression analysis were used to analyze variables that predict the outcome for the tendency to adopt big data security analytics. Tests for normality were performed using stem-and-leaf analysis supported by the graphical box plot from SPSS. By eliminating the outliers in the box plot repeatedly, it is possible to achieve a normally distributed dataset.

Homoscedasticity refers to the assumption of the constant variance of errors across all levels of the selected variables (Fabozzi, Focardi, Rachev, & Arshanapalli, 2014). Data could also exhibit heteroscedasticity where the error terms are not constant across levels of the selected variable (Fabozzi et al., 2014). Homoscedasticity was verified by examining the Durbin-Watson statistic produced by SPSS and by visually examining the scatter plots produced by SPSS. Lack of homoscedasticity can lead to heteroscedasticity, which could lead to distortion and the presence of Type 1 errors.

Threats to Validity

Validity in a quantitative study implies that a study allows correct inferences about the question that it was destined to answer (Field, 2013). Validity is also defined as the degree to which an instrument measures what it is supposed to measure (Nachmias & Nachmias, 2008). Threats to validity need to be identified and addressed since these threats will raise questions about the experimenter’s ability to derive inferences. There are several types of validity pertaining to a quantitative study, such as external validity, internal validity, and construct validity. The SPSS reliability analysis function was used to compute the instrument’s reliability, and Cronbach’s coefficient alpha was used to check threats to validity.

External Validity

External validity consists of ensuring that the results obtained from this study are generalizable beyond the context of this study across time and populations (Druckman, Green, Kuklinski, & Lupia, 2011). Although generalizability is represented in terms of sample size, program design, scale, and some other factors, true external validity provides unbiased estimates of the influence of an intervention on the target population (Orr, 2015). For example, the results obtained in one city may not be typical of the results in another city. External validity provides a true estimate of the effect on the target population. True generalizability is difficult to obtain if the research is always conducted with a convenient and cooperative population, without considering the population of interest (Orr, 2015). As big data security analytics are part of a pervasive and global technology that is purchased and used on the Internet, the results obtained in one state could be similar to the other states within the United States. Although I used convenience sampling in this study, the findings may be generalized to a similar, large population by obtaining an adequate sample size and by using a measurable instrument (see Wright, 2017). However, the results may not be generalizable to other countries due to geographical, cultural, and economic differences.

The disproportion of IT industries in different regions is another external threat. A large population of small businesses in one location can have an effect on the overall results when compared to sparse populations of small businesses. The effect of the population is minimized by having an online survey distributed to different parts of the United States and aggregating the results from the entire country. External validity can be improved by selecting sites and drawing samples that have a reasonable relationship to the target population (Orr, 2015). Hence, by restricting the survey to a sample that is representative of the population, a statistical generalization can be made, and the threats to external validity can be reduced (Polit & Beck, 2010). With the minimization of external threats, the findings from the study can be safely used across the United States.

Internal Validity

Powelson (2012) posited that internal validity assures the truthfulness of the relationships between the variables. When establishing internal validity, it is essential to answer the question of whether the predictor variables alone caused the outcome variable to change (Nachmias & Nachmias, 2008). The factors that weaken internal validity include extrinsic factors, such as biases in selection criteria, and intrinsic factors, such as the history, maturation, experimental mortality, changes in instrumentation, and the process of testing (Nachmias & Nachmias, 2008). This makes internal validity an important tenant in the research field to support multiple and independent replications (Peters & Pereira, 2017). In essence, internal validity ensures that the threats to weaken the researcher’s ability to draw inferences from the data about the population are addressed. For example, the participants with a potential bias to the topic could threaten internal validity (Powelson, 2012). The participants could also threaten internal validity if the participants are affected by what happens around them. Threats to internal validity are of three different types: (a) single group threats, (b) multiple group threats, and (c) social interaction threats.

Internal validity was overcome by using a survey instrument that has been already tested for internal and external validity. Valier et al. (2008) and Powelson (2012) used a standard survey instrument for DOI theory to study the diffusion of open source software and cloud computing technology respectively. The same instrument was adapted to study the diffusion of big data security analytics, by changing only the technology from cloud computing to big data security analytics, thus eliminating the threats to the validity of the instrument. Additionally, correlational research was used instead of causality, thus eliminating the threats to internal validity (Foster, 2017). Therefore, by using statistical methods for data analysis and by cleaning data with the elimination of outliers, the internal validity of this research could be attained.

Construct Validity

Construct validity affirms that the measuring instrument provides support to the selected theoretical framework within which the research is conducted (Nachmias & Nachmias, 2008). The construct validity of this study is attained by the adequacy of the variable definitions and the measures used to perform this study. Powelson's (2012) survey instrument measured the tendency of small businesses to adopt cloud computing by measuring the relationship between the perceived attributes of the innovation and the adoption of the innovation. This instrument was tested and proven for many years in the field of DOI and will be adept at measuring the perceived attributes of innovation and the tendency of small businesses to adopt big data security analytics. This survey instrument is logically and empirically tied to the concepts and assumptions employed (Nachmias & Nachmias, 2008). Hence, construct validity is made possible because of the ability of the survey instrument used historically in measuring the perceived attributes of innovation that are related to the adoption of innovation (Powelson, 2012; Valier et al., 2008).

Ethical Procedures

Ethical procedures and measures systematically resolve ethical dilemmas and ensure that the research has a moral and ethical bearing (Rowley, 2014). Ethical measures are paramount as researchers, universities, and other scholarly practitioners may refer to this study for further research. The ethical measures undertaken includes a consent form, a disclosure of the choice to participate, a disclosure of the ability to terminate survey at any time during the survey, data storage and protection policies of all data collected during the survey, information about storing or encrypting personal, identifiable information, and compliance with Walden University's Institutional Review Board guidelines.

Information about the survey was sent to the participants in the web participant pool provided by Survey Monkey, through an e-mail or text. The participants were qualified through a screening question presented at the beginning of the survey. In addition, the participants electing to participate in the survey had access to the survey hosted at surveymonkey.com using the URL published to each participant through e-mail. As the e-mails were sent from a service provider called Survey Monkey, the service agreement between the service provider and me as shown in Appendix F. In preparation for the survey, the participants were apprised of the scope, purpose, requirements, and confidentiality requirements of this inquiry. Disclosures to the participants included anonymity of participants' data collected during this survey and the ethical requirements of the Walden University. The Institutional Review Board approval number provided for this research is 11-28-18-0305603. Participants had complete information to contact me at any time regarding confidentiality, privacy, or data protection requirements. Additionally, participants had the contact information of a Walden University representative for any inquiry regarding ethical concerns.

Participants in this online survey were given the option to print a copy of the ethics and confidentiality disclosure, which is the letter of informed consent. At the beginning of the survey, the participants were required to accept the electronic agreement and consent administered through the online portal surveymonkey.com. If any of the participants chose to disagree, the survey would be terminated and would not record any personal information about the participant. During the survey, participants had the ability to terminate the survey at any time without being penalized or threatened for not completing the survey. Participants also were able to withdraw participation or cancel the survey at any time during the survey. At the completion of the survey, participants had an opportunity to review their answers before making the final submission. Although there was no monetary compensation, participants were given credit points and the location of the findings of the survey. Participants also gained more knowledge about the big data security analytics because of their participation in this survey.

To provide security for the information, the survey was administered through a secure protocol. Data collected were stored in a secure and confidential location protected by secure authentication procedures. The participants’ personal information or information about their organization was neither collected nor stored during this data collection process. The survey data was analyzed using the IBM SPSS statistical software. Data collected will be kept for a minimum of 5 years to protect the rights of the participants. Data will be deleted after the 5 years to avoid any attempt to plagiarize or manipulate data.

Summary

In this chapter, I presented the research design of the study and the rationale for the selected research design. The purpose of this quantitative correlational study was to increase the adoption of big data security analytics to detect and prevent advanced, persistent threats from malicious resources, by examining the relationship between small business leaders' perceptions of big data security analytics and their adoption. The participants were IT decision makers in small businesses across the United States. I obtained the sampling frame from the Survey Monkey's web participant pool, and a convenience sample was drawn from the sample frame. In this section, I also provided the rationale for quantitative research and the appropriateness of the correlational design for this study. This section also included the procedures for recruitment, participation, and data analysis. Additionally, I explained the instrumentation and operationalization of constructs along with the cross-sectional survey design for data collection. I fully described Powelson's (2012) PreDOI survey instrument featuring five variables depicting perceived attributes of innovation and an outcome variable. I further listed the procedures for protecting data and encrypting secure information. This research abided by the ethical procedures of the Walden University and ensured reliability and validity of the study by addressing external, internal, and construct validity. In Chapter 4, I included the data collection and analysis methods. I also presented the statistical analysis using SPSS to ensure reliability and validity of the data, and to verify the research question hypotheses.

Chapter 4: Results

Introduction

The purpose of this quantitative correlational study was to examine ways to increase the adoption of big data security analytics among small businesses in the United States by examining the relationship between small business leaders' perceptions of big data security analytics attributes and their adoption. The increase in adoption could detect and prevent advanced, persistent threats from malicious resources and improve the confidentiality, integrity, and availability of data among small businesses. For the purpose of this study, small businesses were defined as firms having fewer than 250 employees (Small Business Administration, 2016). I measured the adoption of big data security analytics using the perceived attributes of innovation: compatibility, complexity, observability, relative advantage, and trialability. The PreDOI survey instrument was used to collect the data from web participants by hosting it on the web using Survey Monkey. In this study, I developed the following five research questions with corresponding hypotheses to analyze the relationship between the perceived attributes of innovation and the adoption of big data security analytics:

Research Question 1: To what extent does the perceived attribute of innovation called relative advantage relate to the slow adoption of big data security analytics among small businesses to detect and prevent advanced, persistent threats from malicious sources?

$H_0$1: There is no correlation between relative advantage and the adoption of big data security analytics.

H_a1: There is a positive correlation between relative advantage and the adoption of big data security analytics.

Research Question 2: To what extent does the perceived attribute of innovation called compatibility relate to the slow adoption of big data security analytics among small businesses to detect and prevent advanced, persistent threats from malicious sources?

H_02: There is no correlation between compatibility and the adoption of big data security analytics.

H_a2: There is a positive correlation between compatibility and the adoption of big data security analytics.

Research Question 3: To what extent does the perceived attribute of innovation called complexity relate to the slow adoption of big data security analytics among small businesses to detect and prevent advanced, persistent threats from malicious sources?

H_03: There is no correlation between complexity and the adoption of big data security analytics.

H_a3: There is a negative correlation between complexity and the adoption of big data security analytics.

Research Question 4: To what extent does the perceived attribute of innovation called observability relate to the slow adoption of big data security analytics among small businesses to detect and prevent advanced, persistent threats from malicious sources?

$H_0$4: There is no correlation between observability and the adoption of big data security analytics.

H_a4: There is a positive correlation between observability and the adoption of big data security analytics.

Research Question 5: To what extent does the perceived attribute of innovation called trialability relate to the slow adoption of big data security analytics among small businesses to detect and prevent advanced, persistent threats from malicious sources?

$H_0$5: There is no correlation between trialability and the adoption of big data security analytics.

H_a5: There is a positive correlation between trialability and the adoption of big data security analytics.

In this chapter, I will provide an overview of the recruitment timeframe and the response rates. I will also provide the data screening and cleaning procedures used in the study along with the demographic characteristics. In addition, I will include a discussion of the descriptive and inferential statistics of the study along with the results of the statistical tests, including bivariate analysis, linear regression, and hypothesis testing in this chapter. Finally, I will summarize the findings and provide a transition to Chapter 5.

Data Collection

Time Frame, Recruitment, Response Rates, and Sample Characteristics

I configured the survey on Survey Monkey’s hosting platform and reviewed it for accuracy using the test and preview methods provided by Survey Monkey. The participants were selected from the Survey Monkey's volunteer web participant pool by using Survey Monkey's data collectors. The data collectors enable selecting participants from the Survey Monkey's target audience using filters that I configured on the Survey administration page. The filters applied included:

- Decision making authorities who are part of IT Software;
- Company with employees fewer than 250;
- IT decision makers, whose position included Chief Technology Officer, Chief Information Officer, Management Information System manager, network designer, network manager, project manager, business analyst, and security administrator/analyst;
- Gender including both male and female;
- Age ranging from 18 to 60+ years old;
- Area limited to all regions within the United States; and
- Screening question to qualify participants having big data familiarity.

The period of data collection lasted for 2 weeks, during which 283 participants took the survey. I excluded 28 participants who answered "no" to the screening question and exited the survey without proceeding further. I further eliminated the 25 speeders whose response time was less than 60 seconds yielding 230 responses. After removing the 12 incomplete responses, the total number of completed responses from the survey was 218. Among the 218 responses, I eliminated 18 of them who did not work for small businesses leaving the total count to 200. After further removing eight participants who were straight liners (i.e., those who speed through survey selecting the same option), the final number of valid responses was 192. Using box plot diagrams, 27 outliers were removed, statistically yielding 165 responses, which exceeded my minimal sample size of 115.

As indicated in Table 1, there were more men than women participants. The demographic profile of participants indicated that the majority of the participants were from the age group of 30–44. The participants from the age group from 18–29 almost equaled the adult group from ages 45–60.

Table 1

Demographic Profile of Participants Based on Gender

Gender	Count	%
Male	91	55.2
Female	74	44.8
Total	165	100

Table 2

Demographic Profile of Participants Based on Age Group

Age	Count	%	Cumulative Percent
18-29	28	17.0	17.0
30-44	89	53.9	70.9
45-60	29	17.6	88.5
60+	19	11.5	100.0
Total	165	100.0	

Study Results

Descriptive Statistics

Participant characteristics. Table G1 in the Appendix G shows the descriptive statistics containing participant characteristics, covered by Questions 1 and 2. The majority of the participants, nearly 37%, indicated that their primary responsibility was IT. Table G2 enumerates the participant characteristics by age. Fifty-three percent of the participants were in the 30–44 years old age category. Regarding education level, the majority of the participants, nearly 42%, had a bachelor's degree, while 28.5% had master's degree and only 6.7 % had doctoral-level degrees.

Small business attributes. The descriptive statistics containing small business attributes are in Table H1. Sixty-three firms, representing 38.2 %, were corporations. The descriptive statistics containing small business employee attributes are in Table H2. Thirty-nine participants, representing 23.6%, indicated that their businesses had 50–99 employees. Similarly, another 39 participants, representing 23.6% of the respondents, indicated that their businesses had 100–149 employees. This count was higher than the 28 participants who indicated that their business had 150–199 employees. Finally, 14 participants, representing 8.5% of the respondents, indicated that their businesses had 200–249 employees. The descriptive statistics containing small businesses' industry classification are in Table H3. The majority of the participants belonged to small businesses whose industry type belonged to professional, science, and technical services.

Big data security analytics' awareness. The descriptive statistics containing big data security analytics' awareness are in Appendix I. Nearly 43.6% of the participants had known big data security analytics for the last 4–6 years, while nearly 21.8% of the participants had known big data security analytics for only the last 1–3 years. Of the 165 participants, 120 of them had attended a presentation about big data security analytics. Similarly, of the 165 participants, 131 had read an advertisement about big data security analytics. Finally, of the 165 participants, 127 of them had previously used big data security analytics.

Descriptive statistics about study variables. The mean, standard deviation, and the Cronbach's alpha for each scale used in the study are presented in Table 3. Measurement validity ensures that the scale measures what it is intended to measure (Vaske, Beaman, & Sponarski, 2017). Internal consistency estimates how the individuals respond to the items within a scale. In this study, the perceived attribute of innovation called complexity had a Cronbach's alpha value of .26 and observability had an alpha value of .32, while other attributes of innovation were closer to or above the acceptable alpha value of .65. However, in general, the reliability score is likely to increase with the number of items in the scale (Vaske et al., 2017).

Table 3

Mean, Standard Deviation, and Cronbach's Alpha for Study Variables

Variable	*M*	*SD*	Cronbach's alpha [a]	Number of items
Relative advantage	5.73	0.90	0.92	8
Compatibility	5.70	0.95	0.81	4
Complexity	3.61	0.65	0.26	6
Observability	4.96	0.80	0.32	4
Trialability	5.52	0.93	0.83	5
Adoption	4.29	1.45	0.74	4

Note. Cronbach's alpha scores show high internal consistency for four of six variables.
[a] Reliability: alpha (*n* = 165) > .65 is acceptable, while > .80 desirable.

Statistical Assumptions Evaluation

Using SPSS, I used the Pearson's product-moment correlation to measure the strength of relationship between the perceived attributes of innovation and the adoption of big data security analytics. Before performing statistical tests, I inspected the data using scatter plots and evaluated them for the presence of outliers and missing values. I also eliminated the incomplete responses by downloading only the completed responses from Survey Monkey. I systematically eliminated those who abandoned the survey in the middle, and those who were disqualified based on the screening questions. There were 50 questions in the survey and 283 participants. After applying data cleaning strategies, the number of valid responses was reduced to 192. After eliminating outliers, the final count of participants was 165. The lower bounds and upper bounds identified before eliminating outliers are in Table 4.

Table 4

Outlier Upper and Lower Limits and Extreme Values

Variable [a]	Lower bound	Upper bound	Min	Max
Relative advantage	5.44	5.76	1.00	7.00
Compatibility	5.41	5.74	1.25	7.00
Complexity	3.42	3.67	1.00	5.67
Observability	4.73	5.02	1.25	7.00
Trialability	5.19	5.53	1.00	7.00
Adoption	4.20	4.63	1.00	7.00

Note. [a] *n* = 192. Complexity had more outliers than other variables, CI = 95%.

I used the histograms (see Figure 4) and scatter plots to observe the presence of outliers statistically. A histogram is an accurate representation of the distribution of numerical data. Scatter plots provide a visual representation of the correlation between two variables for a set of data.

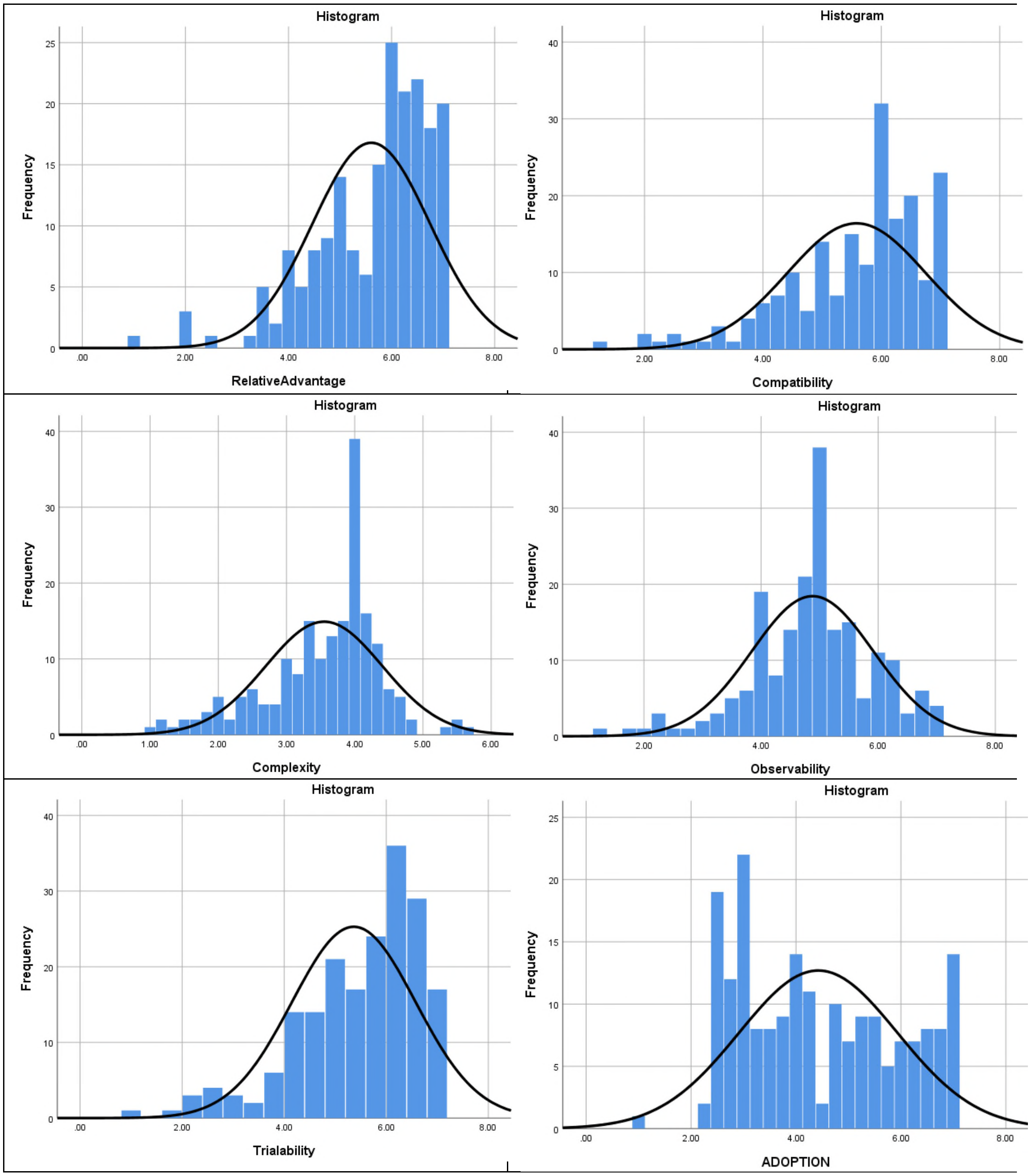


Figure 4. Histograms of the data set.

Finally, I conducted tests of normality using histograms (see Figure 4) and Q-Q plots before removing outliers (see Figure 5) and Q-Q plots after removing outliers (see Figure 6). The histograms relate to one variable. The Q-Q plots compare two probability distributions by plotting their quantiles against each other.

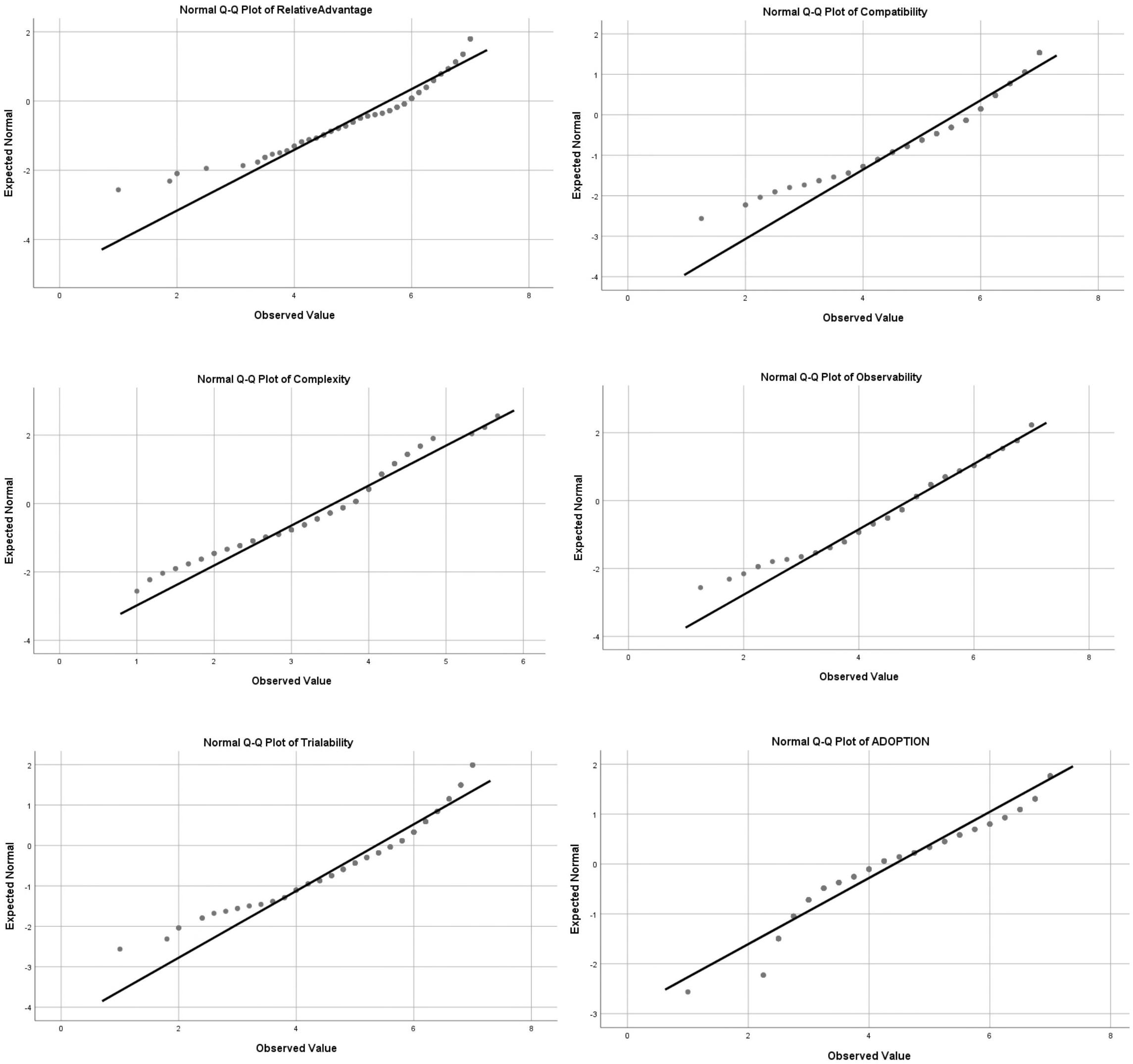


Figure 5. Q-Q plots before removing outliers.

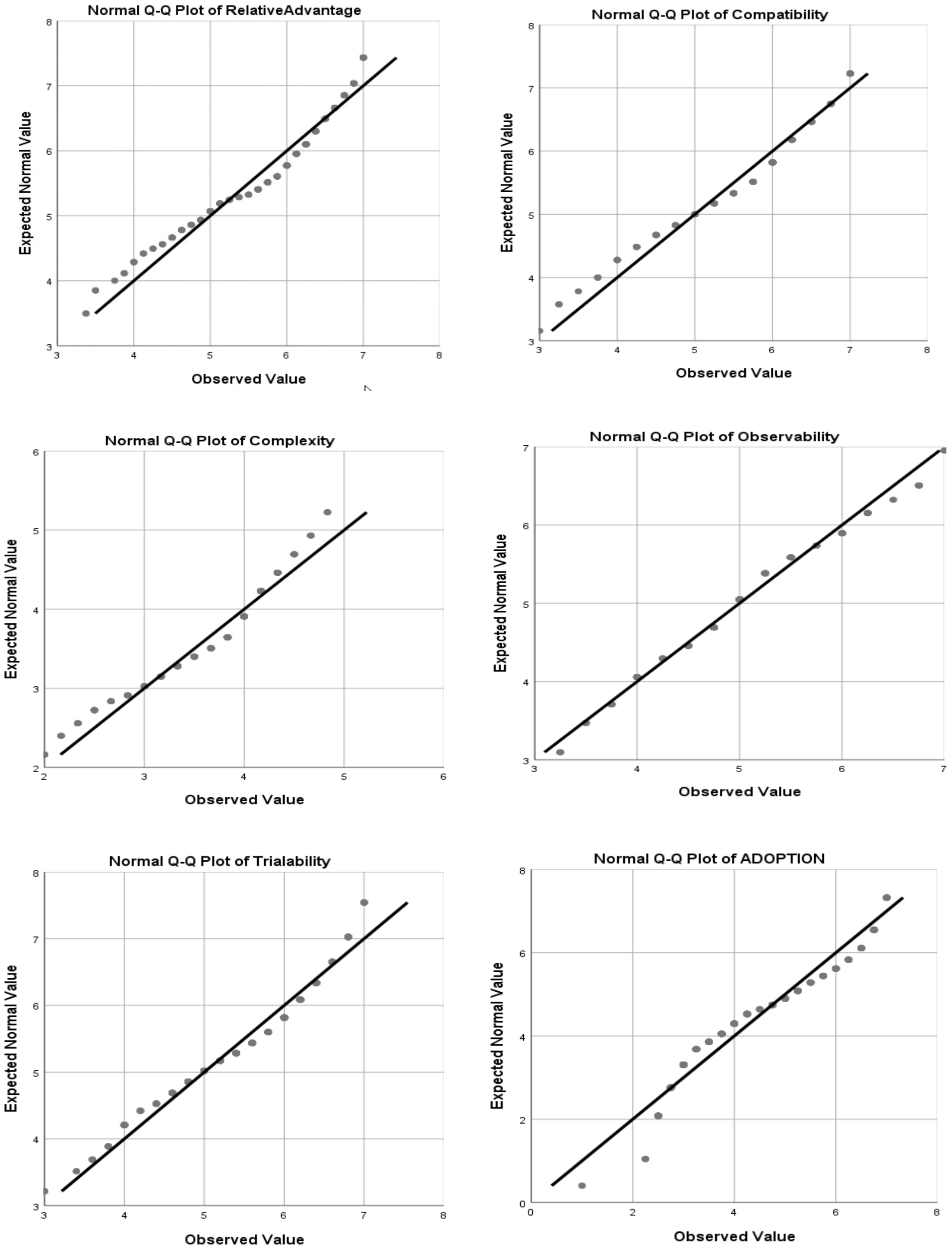


Figure 6. Q-Q plots after removing outliers.

The Q-Q plots appeared to follow a linear pattern and suggested that the data were normally distributed. Q-Q plots can depict the characteristics of a data set. In addition, Q-Q plots are extremely effective in highlighting notable outliers in a data sequence.

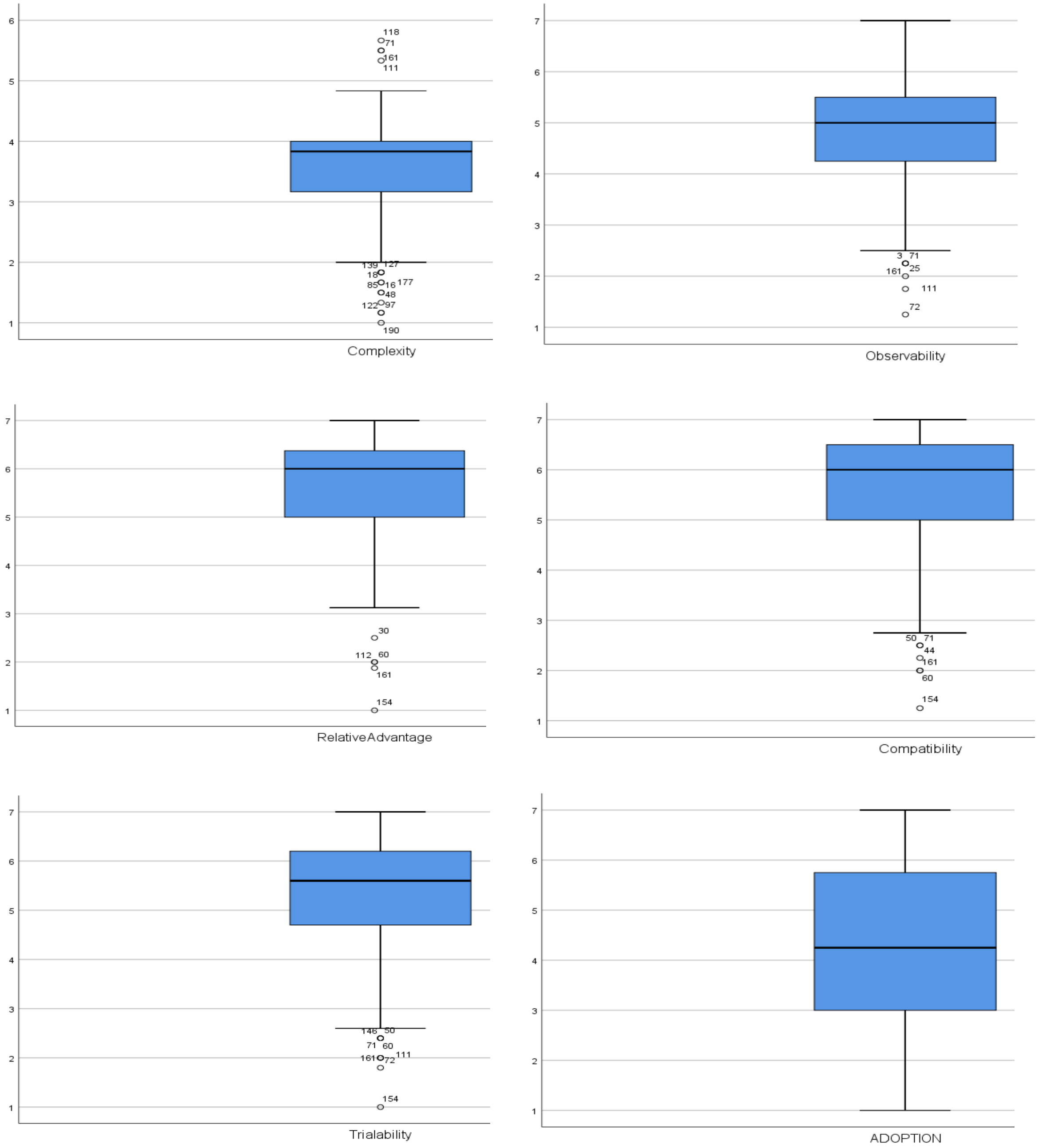


Figure 7. Box plots before removing outliers.

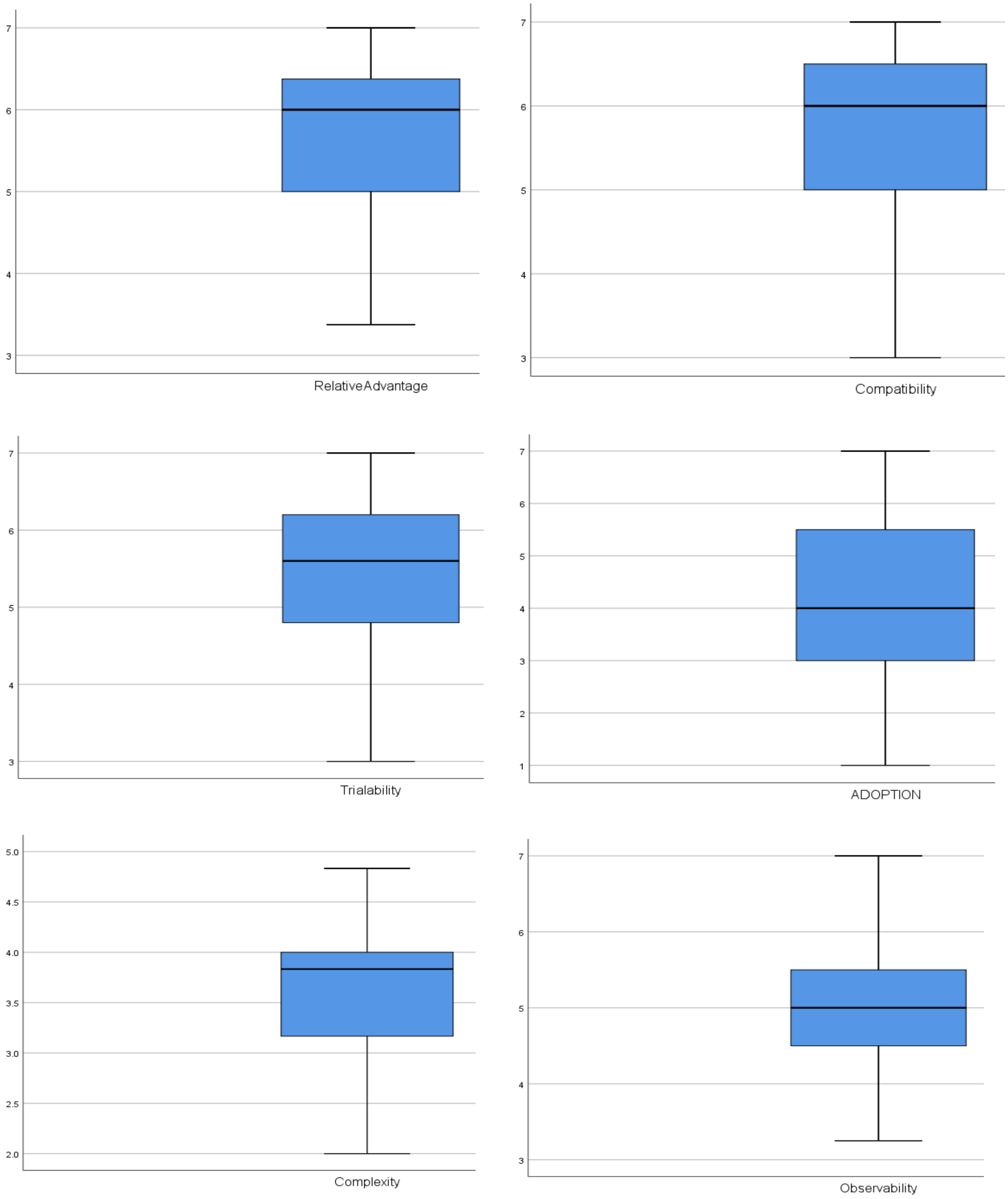


Figure 8. Box plots after removing outliers.

In addition to using Q-Q plots for examining normality, the Shapiro-Wilk test of normality was also included (see Table 5). The Shapiro-Wilk test of normality showed values of *p* less than .05. Hence, the null hypothesis presuming normally distributed data was rejected. Therefore, I concluded that the responses were not from a normally distributed population.

Table 5

Shapiro-Wilk Test of Normality

	Shapiro-Wilk	
Variable	Statistic	Sig.
Relative advantage	.93	.00
Compatibility	.93	.00
Complexity	.94	.00
Observability	.97	.00
Trialability	.95	.00
Adoption	.93	.00

Note. No missing values

Research Question 1 Findings

The research question asked to what extent the perceived attribute of innovation called relative advantage relates to the slow adoption of big data security analytics among small businesses to detect and prevent advanced, persistent threats from malicious sources. The null hypothesis stated that there is no correlation between relative advantage and the adoption of big data security analytics. The first alternate hypothesis stated that there is a positive correlation between the relative advantage and the adoption of big data security analytics. I performed the linear regression analysis to analyze the relationship between the perceived attribute of innovation called relative advantage and the adoption of big data security analytics.

The linear relationship between the relative advantage and the adoption of big data security analytics can be seen in the scatter plot in Figure 9. The regression equation for predicting the adoption of big data security analytics was:

Predicted adoption = .083 * *Relative advantage* + 3.817

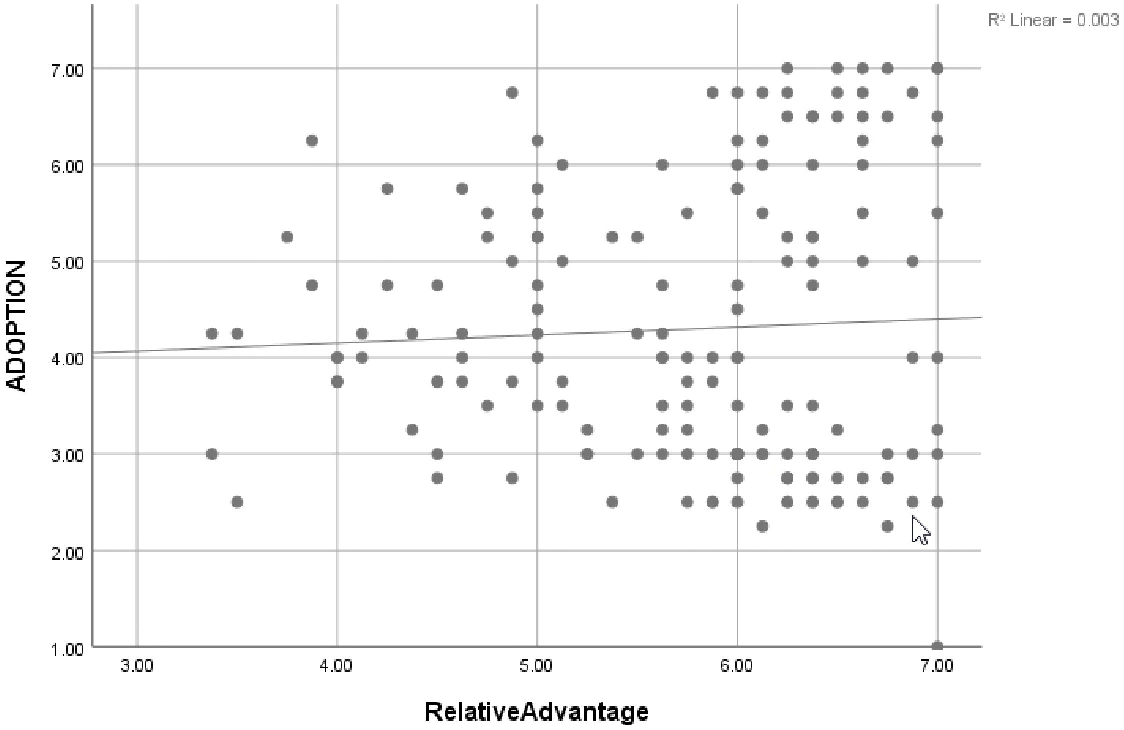


Figure 9. Scatter plot depicting relative advantage and adoption. The best-fit line and R^2 coefficient based on the observed data set.

Results of the Pearson correlation indicated that there was no significant positive association between relative advantage and adoption, *r*(163) = .052, *p* = .507. The results of the Chi-square analysis revealed a nonsignificant positive association between relative advantage and adoption of big data security analytics [χ^2(1, n = 165) = 509, *p* = .93]. Thus, I concluded that there was not a statistically significant relationship between relative advantage and adoption of big data security analytics, and the null hypothesis was true. Approximately 0% of the variance of the adoption was associated with relative advantage. In addition, results of the correlation coefficient for the nonparametric test, Spearman correlation denoted by Spearman's Rho, indicated that there was no significant positive correlation between relative advantage and adoption of big data security analytics, $r_S(163) = .004$, $p = .964$. As hypothesized, the null hypothesis was accepted ($p > .05$) and relative advantage was deemed not significantly related to adoption of big data security analytics as it had only weak positive correlation.

Research Question 2 Findings

The second research question asked to what extent the perceived attribute of innovation called compatibility relates to the slow adoption of big data security analytics among small businesses to detect and prevent advanced, persistent threats from malicious sources. The null hypothesis stated that there is no correlation between compatibility and the adoption of big data security analytics. The second alternate hypothesis stated that there is a positive correlation between compatibility and the adoption of big data security analytics. I performed the linear regression analysis to analyze the relationship between the perceived attribute of innovation called compatibility and the adoption of big data security analytics.

The linear relationship between the compatibility and the adoption of big data security analytics can be seen in the scatter plot in Figure 10. The regression equation for predicting the adoption of big data security analytics was:

$$\textit{Predicted adoption} = .176 * \textit{Compatibility} + 3.289$$

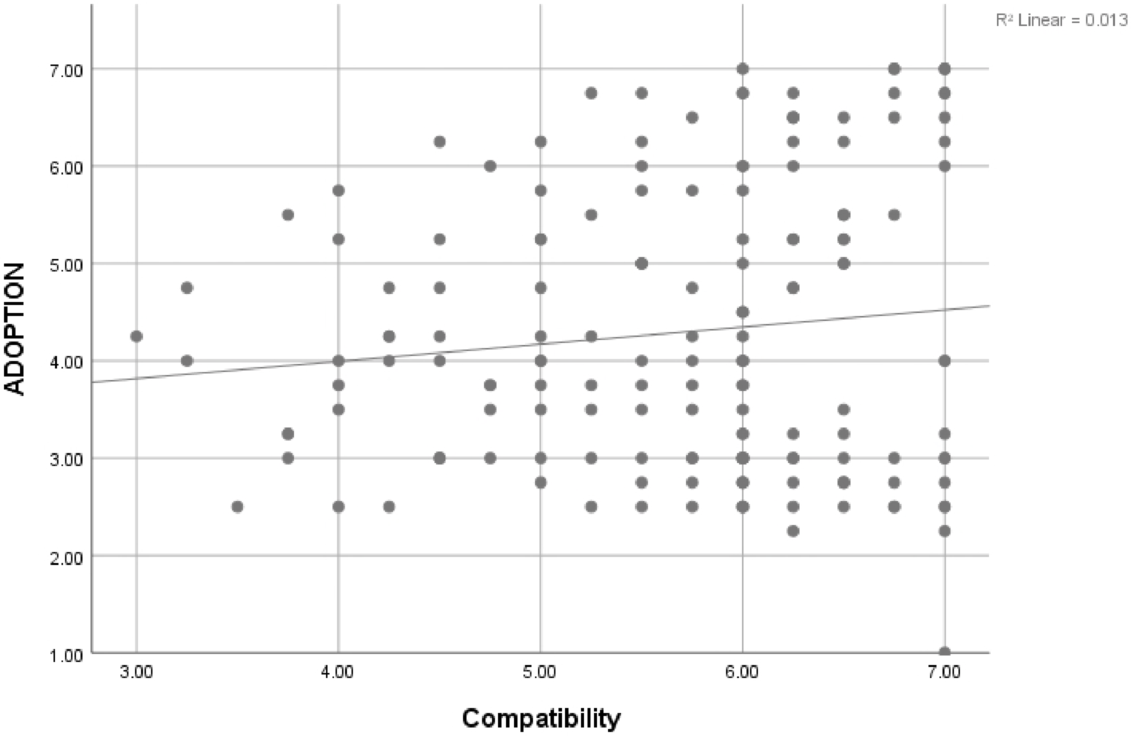


Figure 10. Scatter plot depicting compatibility and adoption. The best-fit line and R^2 coefficient based on the observed data set.

Results of the Pearson correlation indicated that there was no significant positive association between compatibility and adoption, *r*(163) = .116, *p* = .139. The results of the Chi-square analysis revealed a nonsignificant association between compatibility and adoption of big data security analytics [χ^2(1, n = 165) = 303, *p* = .73]. Thus, I concluded that there was not a statistically significant relationship between compatibility and adoption of big data security analytics, and the null hypothesis was true. Approximately 1% of the variance of the adoption was associated with compatibility. In addition, results of the correlation coefficient for the nonparametric test, Spearman correlation denoted by Spearman's Rho, indicated that there was no correlation between compatibility and adoption of big data security analytics, r_s (163) = .068, *p* = .385. As hypothesized, the null hypothesis was accepted ($p > .05$) and compatibility was deemed not significantly related to adoption of big data security analytics as it had only weak positive correlation.

Research Question 3 Findings

The third research question asked to what extent the perceived attribute of innovation called complexity relates to the slow adoption of big data security analytics among small businesses to detect and prevent advanced, persistent threats from malicious sources. The null hypothesis stated that there is no correlation between complexity and the adoption of big data security analytics. The third alternate hypothesis stated that there is a negative correlation between complexity and the adoption of big data security analytics. I performed the linear regression analysis to analyze the relationship between the perceived attribute of innovation called complexity and the adoption of big data security analytics.

The linear relationship between the complexity and the adoption of big data security analytics can be seen in the scatter plot in Figure 11. The regression equation for predicting the adoption of big data security analytics was:

Predicted adoption = -1.068 * *Complexity* + 8.156

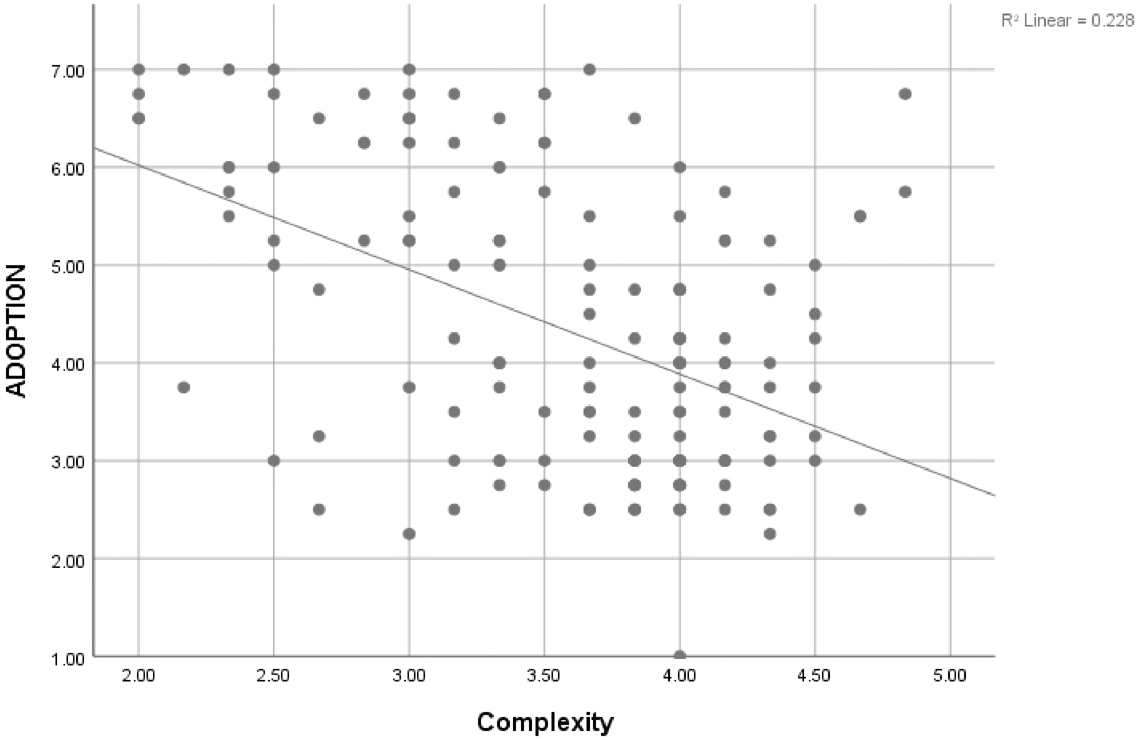


Figure 11. Scatter plot depicting complexity and adoption. The best-fit line and R^2 coefficient based on the observed data set.

Results of the Pearson correlation indicated that there was a significant negative association between complexity and adoption, *r*(163) = -.478, *p* < .01. The results of the Chi-square analysis revealed a significant association between complexity and adoption of big data security analytics [χ^2 (1, n = 165) = 423, *p* = .001]. Thus, I concluded that there was a statistically significant relationship between complexity and adoption of big data security analytics, and I rejected the null hypothesis. The alternative hypothesis, which stated that the lower the complexity, the higher the adoption, was instead accepted. Conversely, the higher the complexity, the lower will be the degree of adoption. Approximately 22% of the variance of the adoption was associated with complexity. In addition, results of the correlation coefficient for the nonparametric test, Spearman correlation denoted by Spearman’s Rho, indicated that there was a negative correlation between complexity and adoption of big data security analytics, $r_s(164) = -.408$, $p < .05$. As hypothesized, the null hypothesis was rejected ($p < .05$) and complexity was deemed negatively correlated to adoption of big data security analytics.

Research Question 4 Findings

The fourth research question asked to what extent the perceived attribute of innovation called observability relates to the slow adoption of big data security analytics among small businesses to detect and prevent advanced, persistent threats from malicious sources. The null hypothesis stated that there is no correlation between observability and the adoption of big data security analytics. The fourth alternate hypothesis stated that there is a positive correlation between observability and the adoption of big data security analytics. I performed the linear regression analysis to analyze the relationship between the perceived attribute of innovation called observability and the adoption of big data security analytics.

The linear relationship between the complexity and the adoption of big data security analytics can be seen in the scatter plot in Figure 12. The regression equation for predicting the adoption of big data security analytics was:

$$\textit{Predicted adoption} = .358 * \textit{Observability} + 2.515$$

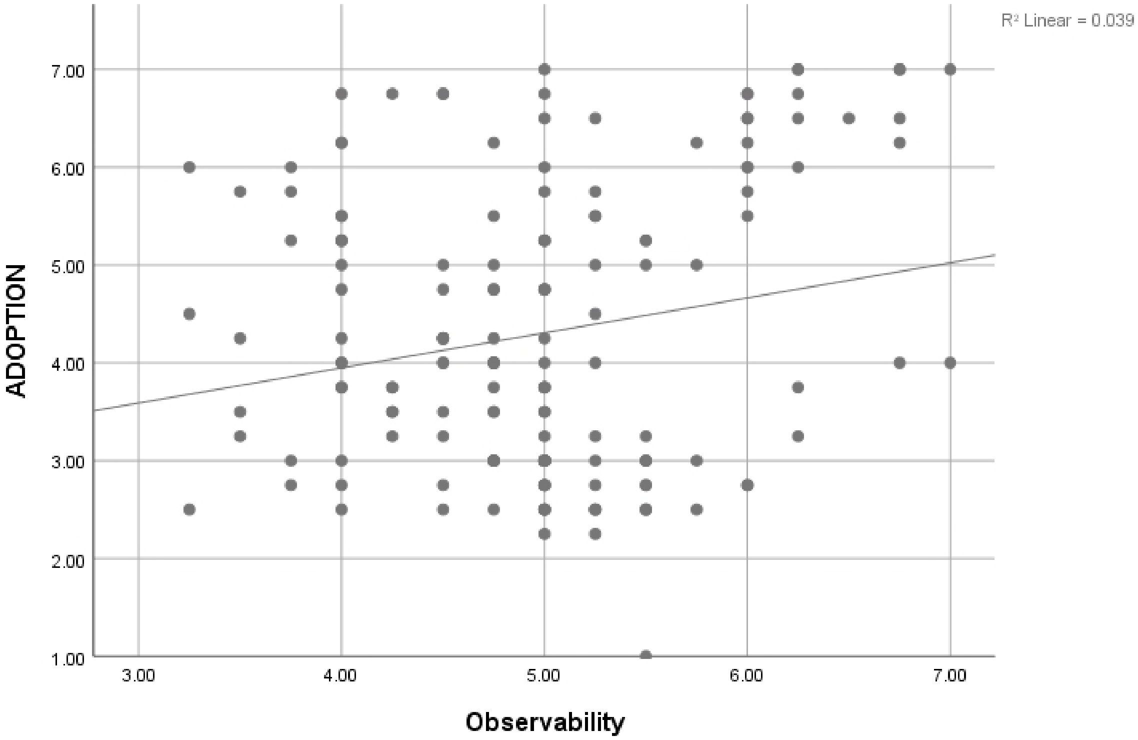


Figure 12. Scatter plot depicting observability and adoption. The best-fit line and R^2 coefficient based on the observed data set.

Results of the Pearson correlation indicated that there was significant positive association between observability and adoption, *r*(163) = .198, *p* = .005. The results of the Chi-square analysis revealed a significant association between observability and adoption of big data security analytics [χ^2 (1, n = 165) = 368, *p* = .004]. Thus, I concluded that there was a statistically significant relationship between observability and adoption of big data security analytics, and I rejected the null hypothesis. The alternative hypothesis, which stated that the higher the observability, the higher the adoption, was instead accepted. Approximately 4% of the variance of the adoption was associated with observability. As hypothesized, the null hypothesis was rejected (*p* < .05) and observability was deemed significantly related to adoption of big data security analytics with positive correlation.

Research Question 5 Findings

The fifth research question asked to what extent the perceived attribute of innovation called trialability relates to the slow adoption of big data security analytics among small businesses to detect and prevent advanced, persistent threats from malicious sources. The null hypothesis stated that there is no correlation between trialability and the adoption of big data security analytics. The fifth alternate hypothesis stated that there is a positive correlation between trialability and the adoption of big data security analytics. I performed the linear regression analysis to analyze the relationship between the perceived attribute of innovation called trialability and the adoption of big data security analytics.

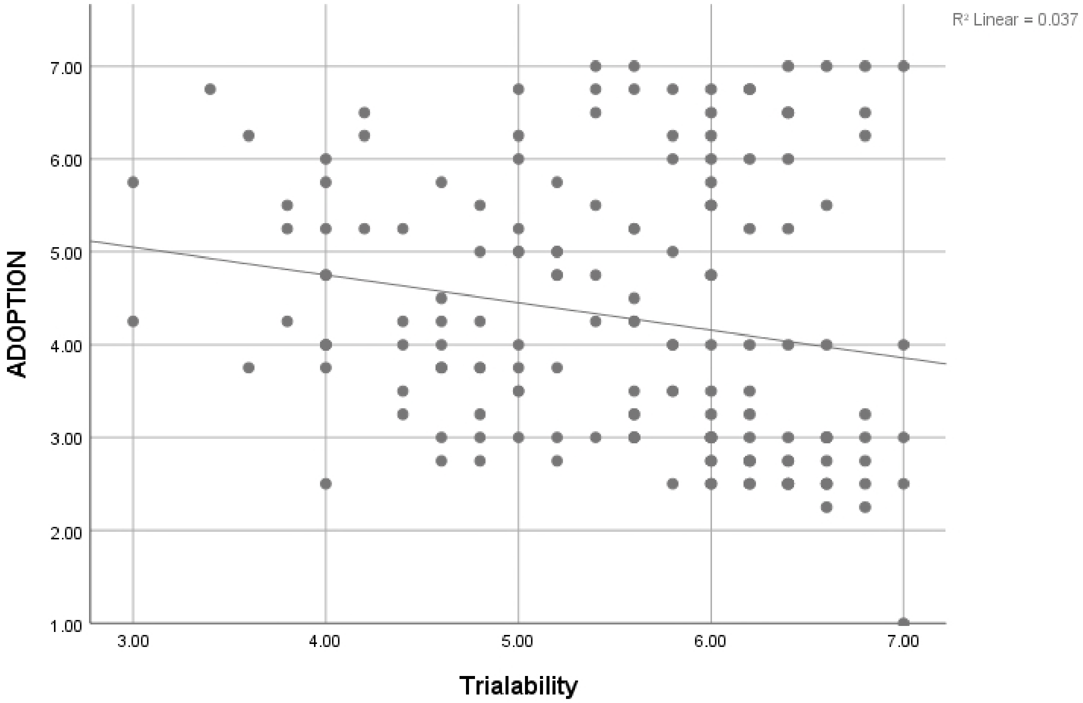


Figure 13. Scatter plot depicting trialability and adoption. The best-fit line and R^2 coefficient based on the observed data set.

The linear relationship between the trialability and the adoption of big data security analytics can be seen in the scatter plot in Figure 13. The regression equation for predicting the adoption of big data security analytics was

Predicted adoption = -.298 * *Trialability* + 5.942

The results of the Pearson correlation indicated that there was significant negative association between trialability and the adoption, $r(163) = -.192$, $p = .014$. The results of the Chi-square analysis revealed a nonsignificant association between trialability and adoption of big data security analytics [χ^2 (1, n = 165) = 415, $p = .101$]. Based on Pearson correlation, I concluded that there was a statistically significant negative relationship between trialability and adoption of big data security analytics, and I rejected the null hypothesis. This negative relationship contradicts the original hypothesis and it could be due to lack of understanding of the trialability of this innovation. Approximately 4% of the variance of the adoption was associated with trialability. In addition, results of the correlation coefficient for nonparametric test, Spearman correlation denoted by Spearman’s Rho, indicated that there was a negative correlation between trialability and the adoption of big data security analytics, r_s (164) = -.253, $p = .001$. As hypothesized, the null hypothesis was rejected ($p < .05$) and trialability was deemed significantly related to adoption of big data security analytics with negative correlation.

Multiple Linear Regression Analysis

I performed multiple regression analysis to analyze the ability of perceived attributes of innovation, such as relative advantage, compatibility, complexity, observability, and trialability to predict the adoption of big data security analytics. An R^2

of .36 (see Table 6) indicated that 36% of the variation of adoption was explained by relative advantage, compatibility, complexity, observability, and trialability. Further, the effects of complexity and trialability significantly predicted the adoption of big data security analytics, while the other variables did not, which can be seen in the regression coefficient table (see Table 7).

Table 6

Model Summary

		R	Change statistics		Durbin-
Model	R	Square	F Change	Sig. F chg.	Watson
1	.602[a]	.362	18.058	.000	1.930

Note. [a] Predictors: Rel. advantage, Compatibility, Complexity, Observability, Trialability.

Table 7

Multivariate Regression Analysis

Independent	Unstandardized . Coefficients		Standard. Coefficient s			95% CI for B	
Variable [a]	*b*	Std. Error	Beta	*t*	Sig.	Lower	Upper
Rel. advantage	.117	.187	.073	.624	.534	-.253	.487
Compatibility	.262	.156	.171	1.682	.095	-.046	.569
Complexity	-.958	.157	-.429	-6.09	.000	-1.26	-.647
Observability	.313	.165	.173	1.902	.059	-.012	.639
Trialability	-.807	.145	-.520	-5.56	.000	-1.094	-.561

Note. [a] Criterion variable: Adoption

Summary

The purpose of Chapter 4 was to analyze the relationship between perceived attributes of innovation, such as relative advantage, compatibility, complexity, observability, and trialability of big data security analytics, and the adoption of big data security analytics. I hypothesized that there would be a positive correlation between the perceived attributes of innovation, such as relative advantage, compatibility, observability, and trialability, and the adoption of big data security analytics. I also hypothesized that there would be a negative correlation between the perceived attribute of innovation called complexity and the adoption of big data security analytics. I cleaned the data using the filters available in Survey Monkey hosting platform and reached a valid sample size of 165, which was greater than the sample size of 115 predicted by G*Power.

A correlation analysis confirmed that there was a weak correlation between the perceived attribute of innovation called relative advantage and the adoption of big data security analytics. Analysis of the perceived attribute of innovation called compatibility confirmed that there was a weak correlation between compatibility and the adoption of big data security analytics. Analysis of the perceived attribute of innovation called complexity proved that there was a negative correlation between complexity and the adoption of big data security analytics. Analysis of the perceived attribute of innovation called observability confirmed that there was a positive correlation between observability and the adoption of big data security analytics. Analysis of the perceived attribute of innovation called trialability confirmed that there was a negative correlation between trialability and the adoption of big data security analytics. I will discuss the interpretations of the research findings, limitations of the study, and my recommendations for further research in Chapter 5. I also will review implications for scholar practitioners and for positive social change.

Chapter 5: Discussion, Conclusions, and Recommendations

Introduction

In this chapter, I will include a general summary, interpretation of findings, the limitations of the study, recommendations for future research, implications for future researchers and positive social change, and a conclusion. The purpose of this quantitative correlational study was to examine ways to increase the adoption of big data security analytics among small businesses in the United States by examining the relationship between small business leaders' perceptions of big data security analytics attributes and their adoption. The increase in adoption could detect and prevent advanced, persistent threats from malicious resources and improve the confidentiality, integrity, and availability of data among small businesses. I based the study on five research questions:

Research Question 1: To what extent does the perceived attribute of innovation called relative advantage relate to the slow adoption of big data security analytics among small businesses to detect and prevent advanced, persistent threats from malicious sources?

Research Question 2: To what extent does the perceived attribute of innovation called compatibility relate to the slow adoption of big data security analytics among small businesses to detect and prevent advanced, persistent threats from malicious sources?

Research Question 3: To what extent does the perceived attribute of innovation called complexity relate to the slow adoption of big data security analytics among small businesses to detect and prevent advanced, persistent threats from malicious sources?

Research Question 4: To what extent does the perceived attribute of innovation called observability relate to the slow adoption of big data security analytics among small businesses to detect and prevent advanced, persistent threats from malicious sources?

Research Question 5: To what extent does the perceived attribute of innovation called trialability relate to the slow adoption of big data security analytics among small businesses to detect and prevent advanced, persistent threats from malicious sources?

Interpretation of Findings

Verizon Enterprise (2018) reported that more than 58% of all data breaches occurred in small businesses, and nearly 68% of the data breaches took months or longer to discover. In another study, Horton (2014) reported that 90% of data breaches affected small businesses. Big data security analytics can help to detect incoming threats using techniques, such as agile decision-making, dynamic detection of both known and previously unknown behaviors, and effective detection of malicious behaviors in real time using multifactor approaches (Marchetti et al., 2016). Although adoption of big data analytics was one of the top priorities of organizations, only 29% of executives reported that they were using big data for predictive analytics (Greengard, 2014). The security analytics' survey results published by the SysAdmin, Audit, Network, and Security

Institute revealed that only 25% of the big data secure analytic solutions monitor threat events and reporting (Shackleford, 2013).

Using Rogers's (2003) PreDOI theory, I analyzed the adoption of big data security analytics among small businesses in the United States using users' perceived attributes of innovation. Rogers asserted that an individual's knowledge about the innovation, perception of the innovation, and the societal factors surrounding the individual plays a significant role in the individual's decision to adopt the innovation. Rogers further affirmed that technology diffusion could be predicted using the PreDOI theoretical constructs.

Relative Advantage

Rogers (2003) posited that the higher the relative advantage of an innovation, the greater its adoption. Big data security analytics have the ability to analyze unstructured data; however, there was no empirical research to indicate the value or relative advantage of big data analytics (Wamba et al., 2015). In this empirical study, based on the sample size of 165 participants obtained from various small businesses in the United States, the correlation analysis revealed that the IT decision makers did not find the big data security analytics as advantageous to their work situation. Hence, the IT decision makers are less likely to adopt big data security analytics based on its relative advantage.

Compatibility

Rogers (2003) posited that the higher the compatibility of an innovation, the greater its adoption. Big data uses current technologies and has become an enabler of improved decision making for enhanced firm performance (Wamba et al., 2015).

However, based on the sample size of 165 participants obtained from various small businesses in the United States, the correlation analysis revealed that the IT decision makers did not find the big data security analytics as compatible with other technologies. Hence, the IT decision makers are less likely to adopt big data security analytics based on its compatibility.

Complexity

Rogers (2003) posited that the lower the complexity of an innovation, the greater its adoption. There is an inherent complexity in processing big data, which contain both structured and unstructured data (Apurva et al., 2017). However, in this empirical study, based on the responses from 165 participants from various small businesses in the United States, the correlation analysis revealed that the IT decision makers did not find the big data security analytics to be complex. Hence, the IT decision makers are more likely to adopt big data security analytics based on its lack of complexity.

Observability

Rogers (2003) posited that the higher the observability of innovation, the greater its adoption. Cloud computing environments, such as the Amazon web services, provide servers, storage, and computation environments to execute big data applications in cloud environments (Feller et al., 2015). In this empirical study, based on the responses from 165 participants from various small businesses in the United States, the correlation analysis revealed that the IT decision makers found the big data security analytics to be observable. Hence, the IT decision makers are more likely to adopt big data security analytics based on its observability.

Trialability

Rogers (2003) posited that the higher the trialability of an innovation, the greater its adoption. Powelson (2012) operationalized trialability to help measure the ability to use an innovation. In this empirical study, based on the responses from 165 participants from various small businesses in the United States, the correlation analysis revealed that there was a negative correlation between the trialability and the adoption of big data security analytics. Hence, the IT decision makers are less likely to adopt big data security analytics based on its current trialability. Although this finding contradicts the original alternative hypothesis, the results could be due to the participants' lack of understanding of trialability.

Limitations of the Study

While the results of this study contribute to the body of literature around relative advantage, compatibility, complexity, observability, and trialability of big data security analytics, there were a few limitations to this study. First, the Cronbach's alpha showed less internal consistency for 2 of the 5 variables. The PreDOI survey instrument devoted four questions to measuring adoption, and only one of them had positive coding, while the remaining three had reverse coding. This could have introduced incorrect results if the participants did not pay attention to the questions that had reverse coding, affecting the true value of the criterion variable called adoption. Second, I collected data through convenience sampling with a self-selection method, which could have presented less accurate results. Third, the sample size was limited to 165. The larger the sample, the better it is for generalization of the study results. Fourth, the study was limited geographically to the small businesses in the United States. The results may not be generalizable to other countries due to economic, ethnic, and cultural differences across different countries. Fifth, far more men than women participated in this study. The results might have been different if more women were included using the gender balancing feature of Survey Monkey. Sixth, the study included people of ages 18 or above. The results could have been different if only particular age groups were included. Seventh, the 4 questions related to adoption were presented towards the end of the survey. The results could have been different if they were mixed with other questions.

Recommendations

Since Verizon Enterprise (2018) reported that more data breaches occurred in small businesses and those data breaches took months or longer to discover, I focused on the small business in the United States. Future researchers can begin to explore mid-size and large businesses that might require big data security analytics to build secure systems and transactions. I limited this study to all regions in the United States. Further research could extend to international regions to understand the adoption of big data security analytics in other parts of the global economy. IT software was a primary filter applied among those used in the selection of the organizations in the web-participant pool. Future research could extend the results of this study and use IT hardware to study the adoption of big data security analytics among IT hardware businesses.

Future research could also include the replication of this study with random sampling to gain more insight into the adoption of big data security analytics. In this study, Cronbach’s alpha indicated that the PreDOI instrument had less internal consistency for 2 of the 5 variables. Future research could include a pilot study to validate the internal consistency of the survey instrument or use another instrument for the same study to study the adoption of big data security analytics among small businesses in the United States. Another study could focus on those small businesses with six to 49 employees.

Only 1 of the 4 questions in the PreDOI related to the adoption of big data security analytics had positive wording in the sentence. Future research could plan to have at least 2 of the 4 questions with positive wording in the sentence. This could help to eliminate an incorrect understanding of the questions by the participants. Also, future research could use the gender balancing and age balancing options in the Survey Monkey hosting platform while fielding the online survey. Finally, future researchers could extend this study to other social organizations, such as nonprofit organizations, so that they can also reap the benefits of adoption of big data security analytics to build secure systems and transactions.

Implications

Big data security analytics is used to analyze structured, semistructured, and unstructured data using cloud-computing technologies. Small businesses are springboards to large businesses, and hence, securing small businesses leads to securing large enterprises. The outcome of this study provided input for positive social change for small businesses, the IT workforce, and for society as a whole.

Positive Social Change for Small Business

The IT decision makers in small businesses consider big data security analytics as a technology that is easy to understand and observe. Based on the findings in this study, the big data security analytics have to become more advantageous and compatible to the current environment so that small business can use them to detect and thwart advanced, persistent threats. Cybersecurity threats, such as spamming, search poisoning, botnets, denial of service, phishing, malware, and website threats, have steadily increased, and data breaches have become a consistently added cost of doing business (Ponemon, 2016). Small businesses can now use the powerful infrastructure of big data security analytics to detect advanced, persistent threats by analyzing the logs collected over a period of time (Farrell, 2016; Li & Oprea, 2016). I will publish the findings of this study in the online Google storage location configured for this study, where participants of the study can access the results and implement the suggestions for their small businesses. The research findings will also be published in the ProQuest Dissertations and Theses database so that more small business entrepreneurs can access them. I will also contact small business forums for further dissemination of the results of the study. By following the recommendations in this study, more small businesses could become increasingly secure by detecting and eliminating against advanced, persistent threats.

Positive Social Change for IT Workforce

The results of this study indicated the areas to improve for building secure transactions and systems. Small businesses are yet to see the relative advantage and compatibility of the big data security analytics. However, the demand for workforce qualifications in big data analytics is increasing. Some organizations have learned to use a few skilled big data practitioners to develop services that encapsulate big data operations (Kim et al., 2015), while others are expecting employees with skills and expertise, to handle large volumes of data used for predictive analytics (Earnshaw et al., 2015). Small businesses can now specialize in big data security analytics to provide more job opportunities to the IT workforce and build secure systems for both small and largescale enterprises, thus bringing significant positive social change into the IT workforce.

Positive Social Change for Society

The focus of this research was to identify ways to increase the adoption of big data security analytics among small businesses in the United States as small businesses are more susceptible to advanced, persistent threats than enterprises (Verizon Enterprise, 2018). Empirical results observed in this study indicated that big data security analytics is less complex, is easy to observe, and trialable. Big data security analytics can intelligently identify undiscovered patterns of attacks and use predictive algorithms to thwart future attacks. In addition, by using big data security analytics, it is possible to detect and eliminate advanced, persistent threats, which the traditional security information and event management tools could not do. The removal of advanced, persistent threats could increase the confidentiality, integrity, and availability of systems and data. Detecting fraudulent transactions and data thefts increases the trust and availability of systems and data. Small businesses could now improve decision making and increase the security of transactions between systems thus bringing positive social change to society by increasing the adoption of big data security analytics.

Conclusions

I examined the relationship of relative advantage, compatibility, complexity, observability, and trialability, and the adoption of big data security analytics. Empirical results indicated that complexity and trialability had a significant negative correlation, observability had a significant positive correlation, while relative advantage and compatibility had a weak positive correlation with the adoption of big data security analytics. To increase the adoption of big data security analytics to detect and thwart advanced, persistent threats from malicious sources; there is a need for demonstrating the relative advantage and compatibility of big data security analytics to the small businesses in the United States. As small businesses act as springboards to larger businesses in the United States, adopting big data security analytics could help to identify and eliminate advanced, persistent threats thus increasing the confidentiality, integrity, and availability of systems and data. Future research could include a study of the adoption of big data security analytics internationally.

References

Abowitz, D. A., & Toole, T. M. (2010). Mixed method research: Fundamental issues of design, validity, and reliability in construction research. *Journal of Construction Engineering & Management, 136*(1), 108-116. doi:10.1061/(ASCE)CO.1943-7862.0000026

Addo-Tenkorang, R., & Helo, P. T. (2016). Big data applications in operations/supply-chain management: A literature review. *Computers & Industrial Engineering, 101*, 528-543. doi:10.1016/j.cie.2016.09.023

Adolph, M. (2014). Big data, its enablers and standards. *PIK – Praxis der Informationsverarbeitung und Kommunikation, 37*(3), 197-204. doi:10.1515/pik-2014-0017

Aminzade, M. (2018). Confidentiality, integrity, and availability – finding a balanced IT framework. *Network Security, 2018*(5). 9-11. doi:10.1016/S1353-4858(18)30043-6

Andreu-Perez, J., Poon, C. C. Y., Merrifield, R. D., Wong, S. T. C., & Yang, G. Z. (2015). Big data for health. *IEEE Journal of Biomedical and Health Informatics, 19*(4), 1193-1208. doi:10.1109/JBHI.2015.2450362

Al-Dhuraibi, Y., Paraiso, F., Djarallah, N., & Merle, P. (2018). Elasticity in cloud computing: State of the art and research challenges. *IEEE Transactions on Services Computing, 11*(2), 430-447. doi:10.1109/TSC.2017.2711009

Alsuhibany, S. (2016). A space-and-time efficient technique for big data security analytics. *Information Technology (Big Data Analysis),* 1-6. doi:10.1109/KACSTIT.2016.7756065

Apurva, A., Ranakoti, P., Yadav, S., Tomer, S., & Roy, N. R. (2017). Redefining cyber security with big data analytics. *2017 International Conference on Computing and Communication Technologies for Smart Nation,* 199-203. doi:10.1109/IC3TSN.2017.8284476

Ashawa, M. (2018). Vulnerability assessment and evaluation of associated attacks on physical and virtual networks. *IUP Journal of Computer Sciences, 12*(2), 43-61. Retrieved from https://www.ijser.org/

Asiedu, E., & Freeman, J. A. (2007). The effect of globalization on the performance of small and medium-sized enterprises in the United States: Does owner's race/ethnicity matter? *American Economic Review, 97*(2), 368-372. doi:10.1257/aer.97.2.368

Azucar, D., Marengo, D., & Settanni, M. (2018). Predicting the big 5 personality traits from digital footprints on social media: A meta-analysis. *Personality and Individual Differences, 124,* 150-159. doi:10.1016/j.paid.2017.12.018

Bardi, M., Xianwei, Z., Shuai, L., & Fuhong, L. (2015). Big data security and privacy: A review. *China Communications, 11*(14), 135-145. doi:10.1109/CC.2014.7085614

Barocas, S., & Nissenbaum, H. (2014). Computing ethics big data's end run around procedural privacy protections. *Communications of the ACM,* 31-33. doi:10.1145/2668897

Basole, C. R., Braunstein, L. M., & Sun, J. (2015). Data and analytics challenges for a learning healthcare system. *Journal of Data and Information Quality, 6*(2-3). doi:10.1145/2755489

Battersby, M. E. (2014). Mitigating risk to protect your business. *Convenience Store Decisions, 25*(10), 116-119. Retrieved from https://cstoredecisions.com/2014/09/30/54431/

Bello, O., & Zeadally, S. (2016). Intelligent device-to-device communication in the Internet of Things. *IEEE Systems Journal, 10*(3), 1172-1182. doi:10.1109/JSYST.2014.2298837

Berghel, H. (1997). Email -- the good, the bad, and the ugly. *Communications of the ACM, 40*(4), 11-15. doi:10.1145/248448.248450

Blazquez, D., & Domenech, J. (2018). Big data sources and methods for social and economic analyses. *Technological Forecasting & Social Change, 130,* 99-113. doi:10.1016/j.techfore.2017.07.027

Bouchard, K., & Giroux, S. (2015). Smart homes and the challenges of data. *Proceedings of the 8th ACM International Conference on Pervasive Technologies Related to Assistive Environments, 66.* doi:10.1145/2769493.2769519

Brewer, R. (2014). Advanced persistent threats: Minimizing the damage. *Network Security, 2014*(4), 5-9. doi:10.1016/S1353-4858(14)70040-6

Caldarola, E. G., Picariello, A., & Castelluccia, D. (2015). Modern enterprises in the bubble: Why big data matters. *ACM SIGSOFT Software Engineering Notes, 40*(1), 1-4. doi:10.1145/2693208.2693228

Cao, L., He, W., Guo, X., & Feng, T. (2016). A scheme for verification on data integrity in mobile multicloud computing environment. *Mathematical Problems in Engineering,* 1-6. doi:10.1155/2016/9267608

Cárdenas, A. A., Manadhata, P. K., & Rajan, S. P. (2013). Big data analytics for security. *IEEE Security & Privacy, 11*(6), 74-76. doi:10.1109/MSP.2013.138

Chen, C., Zhu, X., Shen, P., Hu, J., Guo, S., Tari, Z., & Zomaya, A. Y. (2016). An efficient privacy-preserving ranked keyword search method. *IEEE Transactions on Parallel and Distributed Systems, 27*(4), 951-963. doi:10.1109/TPDS.2015.2425407

Chen, P. C., & Zhang, C.-Y. (2014). Data-intensive applications, challenges, techniques and technologies: A survey on big data. *Information Sciences*, 314-347. doi:10.1016/j.ins.2014.01.015

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

Cui, L., Yu, F. R., & Yan, Q. (2016). When big data meets software-defined networking: SDN for big data and big data for SDN. *IEEE Network, 30*(1), 58-65. doi:10.1109/MNET.2016.7389832

David, C. K. (2014). Viewpoint: Beyond and analysis. *Communications of the ACM, 57*(6), 39-41. doi:10.1145/2602326

De Vet, H. C. W., Mokkink, L. B., Mosmuller, D. G., & Terwee, C. B. (2017). Spearman-Brown prophecy formula and Cronbach’s alpha: Different faces of reliability and opportunities for new applications. *Journal of Clinical Epidemiology, 85,* 45-49. doi:10.1016/j.jclinepi.2017.01.013

Deb, M., & Lomo-David, E. (2014). An empirical examination of customers' adoption of m-banking in India. *Marketing Intelligence & Planning, 32*(4), 475-494. doi:10.1108/MIP-07-2013-0119

Denniston, M. M., Brener, N. D., Kann, L., Eaton, D. K., McManus, T., Kyle, T. M.,… Ross, J. G. (2010). Comparison of paper-and-pencil versus web administration of the Youth Risk Behavior Survey (YRBS): Participation, data quality, and perceived privacy and anonymity. *Computers in Human Behavior, 26*(5), 1054-1060. doi:10.1016/j.chb.2010.03.006

Dodge, H. R., & Robbins, J. E. (1992). An empirical investigation of the organizational life cycle model for small business development and survival. *Journal of Small Business Management, 30*(1), 27-37. Retrieved from https://onlinelibrary.wiley.com/journal/1540627x

Donnelly, R. A. (2007). *The complete idiots guide to statistics* (2nd ed.). New York, NY: The Penguin Group.

Druckman, J. N., Green, D. P., Kuklinski, J. H., & Lupia, A. (2011). *Cambridge handbook of experimental political science.* New York, NY: Cambridge University Press.

Dube, A., & Dube, A. (2016). On the suitability of group lending model in South Sudan's small and medium enterprises sector. *African Review of Economics & Finance, 8*(2), 137-170. Retrieved from https://hdl.handle.net/10520/EJC198162

Earnshaw, R. A., Silva, M. D., & Excell, P. S. (2015). Ten unsolved problems with the Internet of Things. *2015 International Conference on Cyberworlds,* 1-7. doi:10.1109/CW.2015.28

Englmeier, K., & Murtagh, F. (2017). Editorial: What can we expect from data scientists? *Journal of Theoretical & Applied Electronic Commerce Research, 12*(1), 1-5. doi:10.4067/S0718-18762017000100001

Fabozzi, F. J., Focardi, S. M., Rachev, S. T., & Arshanapalli, B. G. (2014). *The business of financial econometrics*. Hoboken, NJ: John Wiley & Sons.

Farrell, R. (2016). Big data security analytics – continuing the innovation wave. *ISSA Journal, 14*(2), 9-34. Retrieved from https://www.issa.org/

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*, 1149-1160. doi:10.3758/BRM.41.4.1149

Feller, E., Ramakrishnan, L., & Morin, C. (2015). Performance and energy efficiency of big data applications in cloud environments: A Hadoop case study. *Journal of Parallel & Distributed Computing, 79*(80), 80-89. doi:10.1016/j.jpdc.2015.01.001

Ferguson, A. G. (2015). Big data and predictive reasonable suspicion. *University of Pennsylvania Law Review, 163*(2), 327-410. doi:10.2139/ssrn.2394683

Field, A. (2013). *Discovering statistics using IBM SPSS Statistics* (4th ed.). Thousand Oaks, CA: Sage.

Foster, T. A. (2017). *Budget planning, budget control, business age, and financial performance in small businesses* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database. (UMI No. 1893566165)

Gahi, Y., Guennoun, M., & Mouftah, H. T. (2016). Big data analytics: Security and privacy challenges. *Computers and Communication,* 952-957. doi:10.1109/ISCC.2016.7543859

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management, 35*(2), 137–144. doi:10.1016/j.ijinfomgt.2014.10.007

Gayadeen, S. M., & Phillips, S. W. (2014). The innovation of community policing and the COPS office: Does diffusion of innovation theory hold in a manipulated environment? *International Journal of Police Science & Management, 16*(3), 228-242. doi:10.1350/ijps.2014.16.3.342

Ghafir, I., Hammoudeh, M., Prenosil, V., Han, L., Hegarty, R., Rabie, K., & Aparicio-Navarro, F. J. (2018). Detection of advanced persistent threat using machine-learning correlation analysis. *Future Generation Computer Systems-The International Journal of Escience, 89*, 249-359. doi:10.1016/j.future.2018.06.055

Ghosh, J. (2016). Big data analytics: A field of opportunities for information systems and technology researchers. *Journal of Global Information Technology Management, 19*(4), 217-222. doi:10.1080/1097198X.2016.1249667

Gil, D., & Song, I. (2016). Modeling and management of big data: Challenges and opportunities. *Future Generation Computer Systems,* 6396-6399. doi:10.1016/j.future.2015.07.019

Gold, S. (2014). Big data and data protection paper from ICO. *Journal of Direct, Data and Digital Marketing Practice, 16,* 137-140. doi:10.1057/dddmp.2014.70

Gomzin, S. (2014). *Hacking point of sale: Payment application secrets, threats, and solutions.* Indianapolis, IN: John Wiley & Sons.

Gray, D. E., & Saunders, M. K. (2016). Beyond survival: How do SMEs win new business and prosper? *Annual International Conference on Business Strategy & Organizational Behaviour*, 11-15. doi:10.5176/2251-1970_BizStrategy16.10

Greengard, S. (2014). Big data = big challenges. *CIO Insight,* 1. Retrieved from https://www.cioinsight.com/blogs/big-data-big-challenges.html/

Gupta, B., Tewari, A., Jain, A., & Agrawal, D. (2017). Fighting against phishing attacks: State of the art and future challenges. *Neural Computing & Applications, 28*(12), 3629-3654. doi:10.1007/s00521-016-2275-y

Gupta, M., & George, J. F. (2016). Toward the development of a big data analytics capability. *Information & Management, 53*(8), 1049-1064. doi:10.1016/j.im.2016.07.004

Han, J., & Liu, J. (2018). Urban spatial interaction analysis using inter-city transport big data: A case study of the Yangtze river delta urban agglomeration of China. *Sustainability, 10*(12), 1-16. doi:10.3390/su10124459

Hansson, S. (2016). Experiments: Why and how? *Science & Engineering Ethics, 22*(3), 613-632. doi:10.1007/s11948-015-9635-3

Harvey, M. (2016). The rise of the LP: The politics of diffusion innovation in the recording industry. *Business History, 58*(7), 1095-1117. doi:10.1080/00076791.2016.1156673

Hashem, I. A., Yaqoob, I., Anuar, N. B., Mokhtar, S., Gani, A., & Khan, S. U. (2015). The rise of 'big data' on cloud computing: Review and open research issues. *Information Systems*, 98-115. doi:10.1016/j.is.2014.07.006

Hashem, T., Datta, S., Islam, T. U., Ali, M. E., Kulik, L., & Tanin, E. (2015). A unified framework for authenticating privacy preserving location based services. *Proceeding GeoRich'15 Second International ACM Workshop on Managing and Mining Enriched Geo-Spatial Data*, 13-18. doi:10.1145/2786006.2786009

Hathaway, M. (Ed.). (2014). *Best practices in computer network defense: Incident detection and response*. Retrieved from https://ebookcentral.proquest.com

Hoque, N., Bhattacharyya, D. K., & Kalita, J. K. (2015). Botnet in DDos attacks: Trends and challenges. *IEEE Communications Surveys and Tutorials, 17*(4), 2242-2270. doi:10.1109/COMST.2015.2457491

Horton, T. (2014). The right technology can help prevent data breaches. *ISO & Agent Weekly, 10*(17), 13-13. Retrieved from https://www.paymentssource.com/iso-agent

Huang, D., Zhao, D., Wei, L., Wang, Z., & Du, Y. (2015). Modeling and analysis in marine big data: Advances and challenges. *Mathematical Problems in Engineering, 2015,* 1-13. doi:10.1155/2015/384742

Humphreys, G. W., & Sui, J. (2015). The salient self: Social saliency effects based on self-bias. *Journal of Cognitive Psychology, 27*(2), 129-140. doi:10.1080/20445911.2014.996156

Hussein, A. A., Hamza, N., & Hefny, H. A. (2013). Attacks on anonymization-based privacy-preserving: A survey for data mining and data publishing. *Journal of Information Security, 4*(2), 101-112. doi:10.4236/jis.2013.42012

Jackson, R. A. (2014). The data behind the curtain. *Internal Auditor, 71*(3), 45-49. Retrieved from https://iaonline.theiia.org/

Jamshidi, D., & Hussin, N. (2016). Islamic credit card adoption understanding: When innovation diffusion theory meets satisfaction and social influence. *Journal of Promotion Management, 22*(6), 897-917. doi:10.1080/10496491.2016.1214206

Jamshidi, D., Hussin, N., & Wan, H. L. (2015). Islamic banking services adoption as a new banking restructure: Examining its adoption from the perspective of DOI theory and trust in Malaysia. *Humanomics, 31*(2), 214-223. doi:10.1108/H-07-2013-0042

Kantarcioglu, M., & Xi, B. (2016). Adversarial data mining: Big data meets cyber security. *CCS '16 Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security,* 1866-1867. doi:10.1145/2976749.2976753

Kaplan, J. M., & Bailey, T., Rezek, C., O’Halloran, D., & Marcus, A. (2015). *Beyond cybersecurity: Protecting your digital business.* Hoboken, NJ: Wiley.

Karim, M. N., Willford, J. C., & Behrend, T. S. (2015). Big data, little individual: Considering the human side of big data. *Industrial & Organizational Psychology, 8*(4), 527-533. doi:10.1017/iop.2015.78

Kaya, Y. (2013). Comparison of quantitative and qualitative research traditions: Epistemological, theoretical, and methodological differences. *European Journal of Education, 48*(2), 311-325. doi:10.1111/ejed.12014

Kim, Y., Kim, Y. H., Lee, G. W., & Huh, E. N. (2015). Survey of bigdata-as-a-service type. *High Performance Computing and Communications,* 1573-1578. doi:10.1109/HPCC-CSS-ICESS.2015.279

Kraemer-Eis, H., & Passaris, G. (2015). SME securitization in Europe. *Journal of Structured Finance, 20*(4), 97-106. Retrieved from http://jsf.iijournals.com/content/20/4/97

Kruidhof, O. (2014). Evolution of national and corporate CERTs – trust, the key factor. In M. Hathway (Ed.), *Best practices in computer network defense: Incident detection and response* (pp. 81–96). doi:10.3233/978-1-61499-372-8-81

Kshetri, N. (2014). Big data’s impact on privacy, security and consumer welfare. *Telecommunications Policy, 38,* 1134-1145. doi:10.1016/j.telpol.2014.10.002

Lambiotte, R., & Kosinski, M. (2014). Tracking the digital footprints of personality. *Proceedings of the IEEE, 102*(12), 1934-1939. doi:10.1109/JPROC.2014.2359054

Li, A., & Oprea, A. (2016). Operational security log analytics for enterprise breach detection. *IEEE Cybersecurity Development,* 15-22. doi:10.1109/SecDev.2016.015

Liang, K., Susilo, W., & Liu, J. (2015). Privacy-preserving ciphertext multi-sharing control for big data storage. *IEEE Transactions on Information Forensics and Security, 10*(8), 1578-1589. doi:10.1109/TIFS.2015.2419186

Lindner, F., & Gaycken, S. (2014). Back to basics: Beyond network hygiene. In M. Hathway (Ed.), *Best practices in computer network defense: Incident detection and response* (pp. 54–64). doi:10.3233/978-1-61499-372-8-54

Lipton, J. D., & Solomon, G. T. (2017). Technology, innovation, entrepreneurship and the small business technology and innovation in small business. *Journal of Small Business, 55*(2), 196-199. doi:10.1111/jsbm.12311

Liu, X. S., Loudermilk, B., & Simpson, T. (2014). Introduction to sample size choice for confidence intervals based on t statistics. *Measurement in Physical Education and Exercise Science, 18*(2), 91-100. doi:10.1080/1091367X.2013.864657

Lund, A., & Lund, M. (2018). Multiple regression analysis using SPSS statistics. Retrieved from https://statistics.laerd.com/spss-tutorials/multiple-regression-using-spss-statistics.php

Magtoto, J., & Roque, A. (2012). Real-time traffic data collection and dissemination from an android smartphone using proportional computation and freesim as a practical transportation system in Metro Manila. *TENCON 2012-2012 IEEE Region 10 Conference,* 1-5. doi:10.1109/TENCON.2012.6412332

Maji, S. K., & Pal, K. (2017). Factors affecting the adoption of e-filing of income tax returns in India: A survey. *IUP Journal of Accounting Research & Audit Practices, 16*(1), 46-66. Retrieved from https://www.iupindia.in/AccountingResearch_AuditPractices.asp

Makar, K. K. (2013). Predict! Teaching statistics using informal statistical inference. *Australian Mathematics Teacher, 69*(4), 34-40. Retrieved from http://www.aamt.edu.au/

Maloshonok, N., & Terentev, E. (2016). The impact of visual design and response formats on data quality in web survey of MOOC students. *Computers in Human Behavior, 62*, 506-515. doi:10.1016/j.chb.2016.04.025

Mannan, S., Nordin, S. M., & Rafik-Galea, S. (2017). Innovation diffusion attributes as predictors to adoption of green fertilizer technology among paddy farmers in Perak state. *Global Business & Management Research*, *9*, 162-170. Retrieved from http://www.irmbrjournal.com/papers/1409660539.pdf

Mansfield-Devine, S. (2016). Securing small and medium-size businesses. *Network Security, 2016*(7), 14-20. doi:10.1016/S1353-4858(16)30070-8

Marchetti, M., Pierazzi, F., Guido, A., & Colajanni, M. (2016). Countering advanced persistent threats through security intelligence and big data analytics. *2016 8th International Conference on Cyber Conflict*, 243-261. doi:10.1109/CYCON.2016.7529438

Marr, B. (2015). *Big data: Using SMART big data, analytics and Metrics to make better decisions and improve performance.* Retrieved from https://ebookcentral.proquest.com

Mazzarol, T., & Clark, D. (2016). The evolution of small business policy in Australia and New Zealand. *Small Enterprise Research*, *23*(3), 239-261. doi:10.1080/13215906.2016.1269242

McCurdy, S. A., & Ross, M. W. (2018). Qualitative data are not just quantitative data with text but data with context: On the dangers of sharing some qualitative data: Comment on Dubois et al. (2018). *Qualitative Psychology, 5*(3), 409-411. doi:10.1037/qup0000088

McMahon, R. (2014). Beyond perimeter defense: Defense-in-depth leveraging upstream security. In M. Hathway (Ed.), *Best practices in computer network defense: Incident detection and response* (pp. 43–53). doi:10.3233/978-1-61499-372-8-43

McNeely, L. C., & Hahm, J. (2014). The big (data) bang: Policy, prospects, and challenges. *Review of Policy Research*, *31*(4), 304-310. doi:10.1111/ropr.12082

Mehmood, A., Natgunanathan, I., Xiang, Y., Hua, G., & Guo, S. (2016). Protection of big data privacy, *IEEE Access*, *4*, 1821-1834. Retrieved from https://ieeexplore.ieee.org/document/7460114

Miao, X., & Zhang, D. (2014). The opportunity and challenge of big data's application in distribution grids. *Electricity Distribution*, *2014 China International Conference,* 962-964. doi:10.1109/CICED.2014.6991847

Miranskyy, A., Hamou-Lhadj, A., Cialini, E., & Larsson, A. (2016). Operational-log analysis for big data systems: Challenges and solutions. *IEEE Software, 33*(2), 52-59. doi:10.1109/MS.2016.33

Mligo, E. S. (2016). *Introduction to research methods and report writing: A practical guide for students and researchers in social sciences and the humanities.* Retrieved from http://ebookcentral.proquest.com

Molina Azorín, J. M., & Cameron, R. (2010). The application of mixed methods in organizational research: A literature review. *The Electronic Journal of Business Research Methods, 8*(2), 95-105. Retrieved from http://www.ejbrm.com/issue/download.html?idArticle=250

Musolesi, M. (2014). Big mobile data mining: Good or evil? *IEEE Internet Computing, 18*(1), 78-81. doi:10.1109/MIC.2014.2

Nachmias, C. F., & Nachmias, D. (2008). *Research methods in social sciences* (7th ed.). New York, NY: Worth.

Nath, N., Hu, Y., & Budge, C. (2016). Information technology and diffusion in the New Zealand public health sector. *Qualitative Research in Accounting and Management, 13*(2), 216-251. doi:10.1108/QRAM-02-2015-0026

Orr, L. L. (2015). 2014 Rossi Award lecture: Beyond internal validity. *Evaluation Review, 39*(2), 167-178. doi:10.1177/0193841X15573659

Pabst, B., Casas, A., & Chinta, R. (2016). Empirical lessons from failure of ERP Systems in small and medium businesses. *Southern Business & Economic Journal, 39*(1), 25-50. Retrieved from http://business.aum.edu/about/southern-business-economic-journal

Perera, C., Ranjan, R., & Wang, L. (2015). End-to-end privacy for open big data markets. *IEEE Cloud Computing, 2*(4), 44-53. doi:10.1109/MCC.2015.78

Peters, S. P., & Pereira, N. (2017). A replication of the internal validity structure of three major teaching rating scales. *Journal of Advanced Academics, 28*(2), 101-119. doi:10.1177/1932202X17701940

Pfleeger, S. L. (2014). The eyes have it: Surveillance and how it evolved. *IEEE Security & Privacy, 12*(4), 74-79. doi:10.1109/MSP.2014.80

Pigni, F., Piccoli, G., & Watson, R. (2016). Digital data streams: Creating value from the real-time flow of big data. *California Management Review, 58*(3), 5-25. doi:10.1525/cmr.2016.58.3.5

Plouffe, C. R., Hulland, J. S., & Vandenbosch, M. (2001). Research report: Richness versus parsimony in modeling technology adoption decisions - understanding merchant adoption of a smart card-based payment system. *Information Systems Research, 12*(2), 208-222. doi:10.1287/isre.12.2.208.9697

Polit, D., & Beck, C. (2010). Generalization in quantitative and qualitative research: Myths and strategies. *International Journal of Nursing Studies, 47*(11), 1451-1458. doi:10.1016/j.ijnurstu.2010.06.004

Ponemon, L. (2015, May 15). 2015 Cost of data breach study. Retrieved from http://www.jmco.com/media/Ponemon-Data-Beach-2015-Report.pdf

Ponemon, L. (2016, June 15). Cost of a data breach 2016. Retrieved from https://securityintelligence.com/cost-of-a-data-breach-2016/

Powelson, S. E. (2012). *An examination of small businesses' propensity to adopt cloud-computing innovation* (Doctoral dissertation). Retrieved from Dissertations & Theses @ Walden University. (UMI No. 963525817).

Prescott, E. M. (2014). Big data and competitive advantage at Nielsen. *Management Decision, 52*(3), 573-601. doi:10.1108/MD-09-2013-0437

Prouska, R., Psychogios, A. G., & Rexhepi, Y. (2016). Rewarding employees in turbulent economies for improved organisational performance. *Personnel Review, 45*(6), 1259-1280. doi:10.1108/PR-02-2015-002

Puri, C., & Dukatz, C. (2015). Analyzing and predicting security event anomalies: Lessons learned from a large enterprise big data streaming analytics deployment. *Database and Expert Systems Applications*, 152-158. doi:10.1109/DEXA.2015.46

Qiu, J., Wu, Q., Ding, G., Xu, Y., & Feng, S. (2016). A survey of machine learning for big data processing. *EURASIP Journal on Advances in Signal Processing, 2016*(1), 1-16. doi:10.1186/s13634-016-0355-x

Rahman, N., & Aldhaban, F. (2015). Assessing the effectiveness of big data initiatives. *2015 Proceedings of PICMET'15: Management of the Technology Age,* 478-484. doi:10.1109/PICMET.2015.7273189

Rassam, M. A., Maarof, M. A., & Zainal, A. (2017). Big data analytics adoption for cybersecurity: A review of current solutions, requirements, challenges and trends.

Journal of Information Assurance & Security, 12(4), 124-145. Retrieved from http://www.mirlabs.org/jias/index.html

Richards, K. (2013). Big data analytics: New patterns emerge for security. *Information Security, 15*(5), 18-24. Retrieved from https://www.infosecurity-magazine.com/

Rigoni, R., & Lindstrom, G. (2014). Computer network defense: New threats and trends. In M. Hathway (Ed.), *Best practices in computer network defense: Incident detection and response* (pp. 19–29). doi:10.3233/978-1-61499-372-8-19

Rogers, A. D. (2016). *Examining small business adoption of computerized accounting systems using the technology acceptance model* (Doctoral dissertation). Retrieved from Dissertations & Theses @ Walden University. (UMI No. 1758891604).

Rogers, E. M. (2003). *Diffusion of innovations* (5th ed.). New York, NY: The Free Press.

Rosenthal, M. (2016). Qualitative research methods: Why, when and how to conduct interviews and focus groups in pharmacy research. *Currents in Pharmacy Teaching & Learning, 8*(4), 509-516. doi:10.1016/j.cptl.2016.03.021

Rowley, H. (2014). Going beyond procedure: Engaging with the ethical complexities of being an embedded researcher. *Management in Education, 28*(1), 19-24. doi:10.1177/0892020613510119

Rubinstein, I. S. (2013). Big data: The end of privacy or a new beginning? *International Data Privacy Law, 3*(2), 74-87. doi:10.1093/idpl/ips036

Saggi, M. K., & Jain, S. (2018). A survey towards an integration of big data analytics to big insights for value-creation. *Information Processing & Management, 54*(5), 758-790. doi:10.1016/j.ipm.2018.01.010

Samar, R., Ghani, M. A., & Alnaser, F. M. (2017). Predicting customer's intentions to use internet banking: The role of technology acceptance model (TAM) in e-banking. *Management Science Letters, 7*(11), 513-524. doi:10.5267/j.msl.2017.8.004

Sănchez-Fernăndez, J., Mućoz-Leiva, F., & Montoro-Rīos, F. (2012). Improving retention rate and response quality in web-based surveys. *Computers in Human Behavior, 28*(2), 507-514. doi:10.1016/j.chb.2011.10.023

Sapoetra, J. (2017). Listening, grammar, and reading comprehension skills of the test of English as foreign language: A correlational study. *Humaniora, 8*(1). doi:10.21512/humaniora.v8i1.3692

Satyanarayanan, M., Simoens, P., Xiao, Y., Pillai, P., Chen, Z., Ha, K.,…Amos, B. (2015). Edge analytics in the internet of things. *IEEE Pervasive Computing, 14*(2), 24-31. doi:10.1109/MPRV.2015.32

Schoonenboom, J. (2017). The realist survey: How respondents' voices can be used to test and revise correlational models. *Journal of Mixed Methods Research. 11*(3), 308-327. doi:10.1177/1558689815610997

Sen, D., Ozturk, M., & Vayvay, O. (2016). An overview of big data for growth in SMEs. *Procedia – Social and Behavioral Sciences, 235,* 159-167. doi:10.1016/j.sbspro.2016.11.011

Shackleford, D. (2013, September). SANS security analytics survey. Retrieved from https://www.sans.org/reading-room/whitepapers/analyst/security-analytics-survey-34980

Shamsi, J. A., & Khojaye, M. A. (2018). Understanding privacy violations in big data systems. *IT Professional*, *20*(3), 73-81. doi:10.1109/MITP.2018.032501750

Shohei, H. (2016). Leveraging Torihiki-Jisseki through Japanese small and medium-sized enterprises's overseas businesses. *Annals of Business Administrative Science, 15*(5), 211-220. doi:10.7880/abas.0160731a

Shostack, A. (2014). *Threat modeling: Designing for security.* Retrieved from https://ebookcentral.proquest.com

Small Business Administration. (2016). *Small business profile.* Retrieved from https://www.sba.gov/sites/default/files/files/Size_Standards_Table.pdf

Small Business Administration. (2018). *Small business profile.* Retrieved from https://www.sba.gov/sites/default/files/advocacy/2018-Small-Business-Profiles-US.pdf

Spielgelhalter, D. (2014). The future lies in uncertainty. *Science*, *345*(6194), 264-265. doi:10.1126/science.1251122

Stewart, J. N. (2014). Advanced technologies/tactics, techniques, procedures: Closing the attack window, and thresholds for reporting and containment. In M. Hathway (Ed.), *Best practices in computer network defense: Incident detection and response* (pp. 30–42). doi:10.3233/978-1-61499-372-8-30

Strong, C. (2014). The challenge of big data: What does it mean for the qualitative research industry? *Qualitative Market Research: An International Journal*, *17*(4), 336-342. doi:10.1108/QMR-10-2013-0076

Sutikno, T., Stiawan, D., & Ibnu Subroto, I. M. (2014). Fortifying big data infrastructures to face security and privacy issues. *Telkomnika, 12*(4), 751–752. doi:10.12928/TELKOMNIKA.v12i4.957

Tallon, P. P., Ramirez, R., & 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. doi:10.2753/MIS0742-1222300306

Tamhane, D. S., & Sayyad, S. N. (2015). Big data analysis using HACE theorem. *International Journal of Advanced Research in Computer Engineering & Technology, 4*(1), 18-23. Retrieved from http://ijarcet.org/wp-content/uploads/IJARCET-VOL-4-ISSUE-1-18-23.pdf

Tan, K. H., Zhan, Y., Ji, G., Ye, F., & Chang, C. (2015). Harvesting big data to enhance supply chain innovation capabilities: An analytic infrastructure based on deduction graph. *International Journal of Production Economics, 165,* 223-233. doi:10.1016/j.ijpe.2014.12.034

Tari, Z. (2014). Security and privacy in cloud computing. *IEEE Cloud Computing, 1*(1), 54-57. doi:10.1109/MCC.2014.20

Tan, Z., Nagar, U. T., He, X., Nanda, P., Liu, R. P., Wang, S., … Hu, J. (2014). Enhancing big data security with collaborative intrusion detection. *IEEE Cloud Computing, 1*(3), 27-33. doi:10.1109/MCC.2014.53

Trantopoulos, K., Krogh, V. G., Wallin, M. W., & Woerter, M. (2017). External knowledge and information technology: Implications for process innovation performance. *MIS Quarterly, 41*(1), 287-A8. Retrieved from http://www.misq.org/skin/frontend/default/misq/pdf/appendices/2017/V41I1Appendices/15_13362_RNSI_TrantopoulosAppendices.pdf

Trespalacios, J. J., & Perkins, R. R. (2016). Effects of personalization and invitation email length on web-based survey response rates. *Techtrends: Linking Research & Practice to Improve Learning, 60*(4), 330-335. doi:10.1007/s11528-016-0058-z

Trifu, M. R., & Ivan, M. L. (2014). Big data: Present and future. *Database Systems Journal, 5*(1), 32-41. Retrieved from http://www.scce.ac.in/e_journals/15_4.pdf

Truong, H., Bui, D., & Tran, V. (2015). GSPInsights: Towards an efficient framework for storing and mining massive vehicle location data. *Proceedings of the Sixth International Symposium on Information and Communication Technology,* 25-31. doi:10.1145/2833258.2833282

Uprichard, E. (2013). Sampling: Bridging probability and non-probability designs. *International Journal of Social Research Methodology, 16*(1), 1-11. doi:10.1080/13645579.2011.633391

Vaidya, J., Shafiq, B., Fan, W., Mehmood, D., & Lorenzi, D. (2014). A random decision tree framework for privacy-preserving data mining. *IEEE Transactions on Dependable and Secure Computing, 11*(5), 399-411. doi:10.1109/TDSC.2013.43

Valier, F. M., McCarthy, R. V., & Aronson, J. R. (2008). A primary study of attributes of innovations during the prediffusion stage. *Journal of International Technology and Information Management, 17*(3), 219-234. Retrieved from https://scholarworks.lib.csusb.edu/jitim/vol17/iss3/4

Van De Pas, J., & Van Bussel, G. (2015). ‘Privacy lost – and found?’ The information value chain as a model to meet citizen’s concerns. *Electronic Journal of Information Systems Evaluation, 18*(2), 185-195. Retrieved from http://www.ejise.com/issue/download.html?idArticle=987

Vaske, J. J., Beaman, J., & Sponarski, C. C. (2017). Rethinking internal consistency in Cronbach’s alpha. *Leisure Sciences*, *29*(2), 163-173. doi:10.1080/01490400.2015.1127189

Vayena, E., Salathé, M., Madoff, L. C., & Brownstein, J. S. (2015). Ethical challenges of big data in public health. *PLoS Computational Biology, 11*(2). doi:10.1371/journal.pcbi.1003904

Vera-Baquero, A., Colomo-Palacios, R., & Molloy, O. (2016). Real-time business activity monitoring and analysis of process performance on big-data domains. *Telematics and Informatics*, *33*(3), 793-807. doi:10.1016/j.tele.2015.12.005

Verizon Enterprise. (2018, April 18). 2018 Data break investigations report. Retrieved from http://www.verizonenterprise.com/resources/reports/rp_DBIR_2018_Report_en_xg.pdf

Verma, S. (2017). The adoption of big data services by manufacturing firms: An empirical investigation in India. *Journal of Information Systems and Technology Management*, *14*(1), 39-68. doi:10.4301/S1807-17752017000100003

Viceconti, M., Hunger, P., & Hose, R. (2015). Big data, big knowledge: Big data for personalized healthcare. *IEEE Journal of Biomedical and Health Informatics. 19*(4). doi:10.1109/JBHI.2015.2406883

Vona, L. W. (2016). *Fraud data analytics methodology: The fraud scenario approach to uncovering fraud in core business systems.* Retrieved from https://ebookcentral.proquest.com

Wamba, F. S., Akter, S., Edwards, A., Chopin, G., & Gnanzou, D. (2015). How ‘big data’ can make big impact: Findings from a systematic review and a longitudinal case study. *International Journal of Production Economics, 165,* 234-246. doi:10.1016/j.ijpe.2014.12.031

Waring, T. S., & Alexander, M. (2015). Innovations in inpatient flow and bed management: An action research project in a UK acute care hospital. *International Journal of Operations & Production Management, 35*(5), 751-781. doi:10.1108/IJOPM-06-2013-0275

Washington, A. L. (2014). Government information policy in the era of big data. *Review of Policy Research, 31*(4), 319-325. doi:10.1111/ropr.12081

Waterman, K. K., & Bruening, P. (2014). Big data analytics: Risks and responsibilities. *International Data Privacy Law, 4*(2), 89-95. doi:10.1093/idpl/ipu002

White, P., & Brenkenridge, R. S. (2014). Trade-offs, limitations, and promises of big data in social science research. *Business Source Complete, 31*(4), 331-338. doi:10.1111/ropr.12078

Wilkes, S. (2012). Some impacts of “big data” on usability practice. *Communication Design Quarterly Review, 13*(2), 25-32. doi:10.1145/2424840.2424842

Williams, P., Jr. (2015). *An assessment of a middle school teachers' efforts to integrate technology effectively* (Order No. 3719527). Retrieved from Dissertations & Theses @ Walden University. (UMI No. 1719138133).

Williamson, A. (2014). Big data and the implications for government. *Legal Information Management*, *14*(4), 253-257. doi:10.1017/S1472669614000553

Wright, B. K. (2017). Researching Internet-based applications: Advantages and disadvantages of online survey research, online questionnaire authoring software packages, and web survey services. *Journal of Computer-Mediated Communication, 10*(3). doi:10.1111/j.1083-6101.2005.tb00259.x

Wu, D., Yang, B., Wang, H., Wang, C., & Wang, R. (2016). Privacy-preserving multimedia big data aggregation in large-scale wireless sensor networks. *ACM Transactions on Multimedia Computing Communications and Applications, 12*(4). doi:10.1145/2978570

Wu, X., Zhu, X., Wu, G. Q., & Ding, W. (2013). Data mining with big data. *IEEE Transactions on Knowledge and Data Engineering, 26*(1), 97-107. doi:10.1109/TKDE.2013.109

Xu, L., Jiang, C., Chen, Y., Ren, Y., & Liu, K. (2015). Privacy or utility in data collection? A contract theoretic approach. *IEEE Journal of Selected Topics in Signal Processing, 9*(7), 1256-1269. doi:10.1109/JSTSP.2015.2425798

Xu, L., Jiang, C., Wang, J., Yuan, J., & Ren, Y. (2014). Information security in big data: Privacy and data mining. *IEEE Access, 2*(0), 1149-1176. doi:10.1109/ACCESS.2014.2362522

Yilmaz, K. (2013). Comparison of quantitative and qualitative research traditions: Epistemological, theoretical, and methodological differences. *European Journal of Education, 48*, 311–325. doi:10.1111/ejed.12014

Yong Suk, L. (2017). Entrepreneurship, small businesses and economic growth in cities. *Journal of Economic Geography, 17*(2), 311-343. doi:10.1093/jeg/lbw021

Young, S. (2015). A big data approach to HIV epidemiology and prevention. *Preventive Medicine, 70,* 17-18. doi:10.1016/j.ypmed.2014.11.002

Zhan, Y., & Tan, K. H. (2018). An analytic infrastructure for harvesting big data to enhance supply chain performance. *European Journal of Operational Research.* doi:10.1016/j.ejor.2018.09.018

Zhang, C., Shen, X., Pei, X., & Yao, Y. (2016). Applying big data analytics into network security: Challenges, techniques and outlooks. *Smart Cloud,* 1-6. doi:10.1109/SmartCloud.2016.62

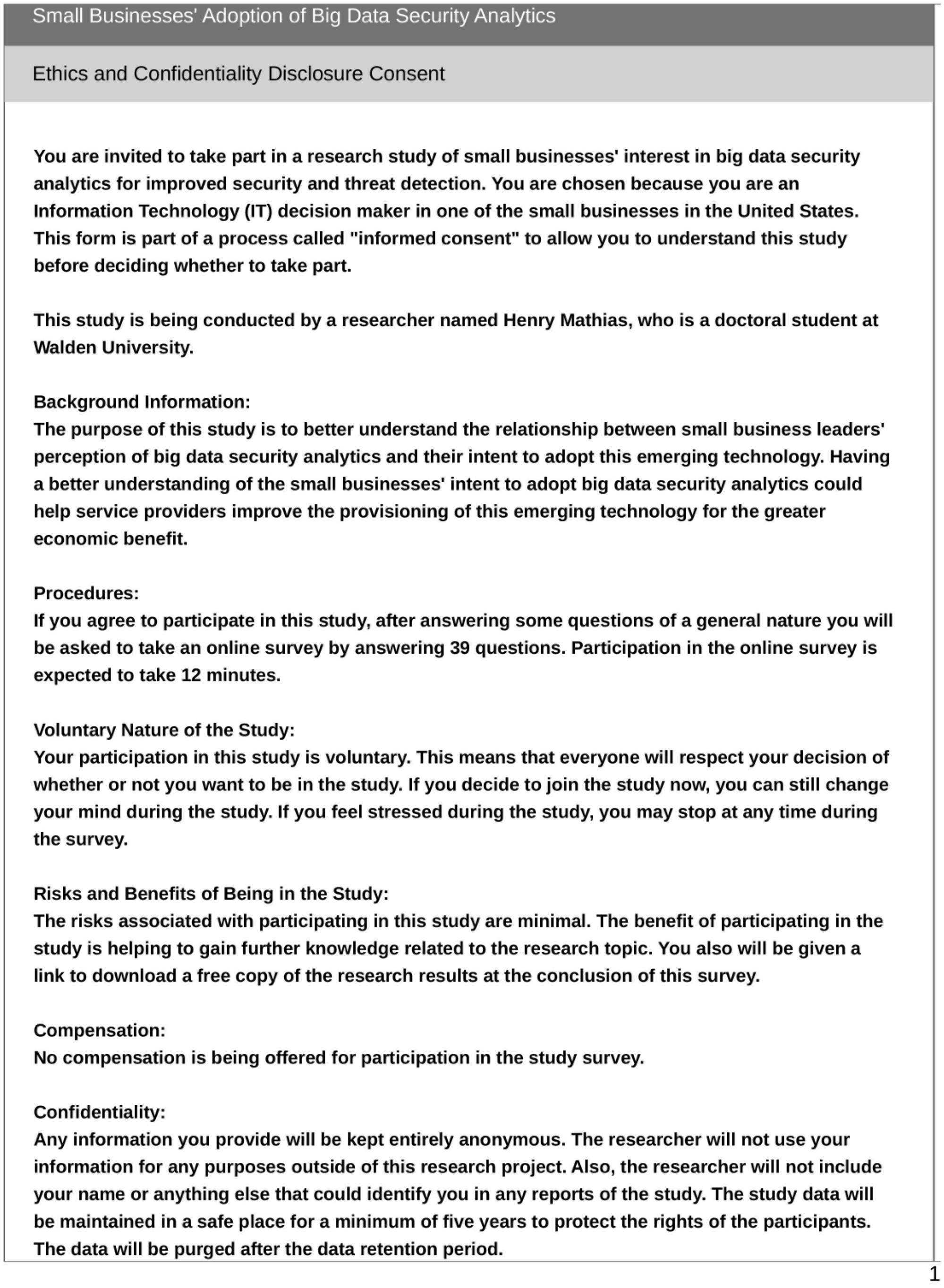
Zhang, H., Zhang, Q., Zhou, Z., Du, X., Yu, W., & Guizani, M. (2015). Processing geo-dispersed big data in an advanced MapReduce framework. *IEEE Network, 29*(5), 24-30. doi:10.1109/MNET.2015.7293301

Zhang, X., Yang, L., Liu, C., & Chen, J. (2013). A scalable two-phase top-down specialization approach for data anonymization using MapReduce on cloud. *IEEE*

Transactions on Parallel and Distributed Systems, 25(2), 363-373. doi:10.1109/TPDS.2013.48

Zhu, T., Xiao, S., Zhang, Q., Gu, Y., Yi, P., & Li, Y. (2015). Emergent technologies in big data sensing: A survey. *International Journal of Distributed Sensor Networks, 2015,* 1-13. doi:10.1155/2015/902982

Appendix A: Sample PreDOI Survey Instrument



Small Businesses' Adoption of Big Data Security Analytics

Ethics and Confidentiality Disclosure Consent

You are invited to take part in a research study of small businesses' interest in big data security analytics for improved security and threat detection. You are chosen because you are an Information Technology (IT) decision maker in one of the small businesses in the United States. This form is part of a process called "informed consent" to allow you to understand this study before deciding whether to take part.

This study is being conducted by a researcher named Henry Mathias, who is a doctoral student at Walden University.

Background Information:
The purpose of this study is to better understand the relationship between small business leaders' perception of big data security analytics and their intent to adopt this emerging technology. Having a better understanding of the small businesses' intent to adopt big data security analytics could help service providers improve the provisioning of this emerging technology for the greater economic benefit.

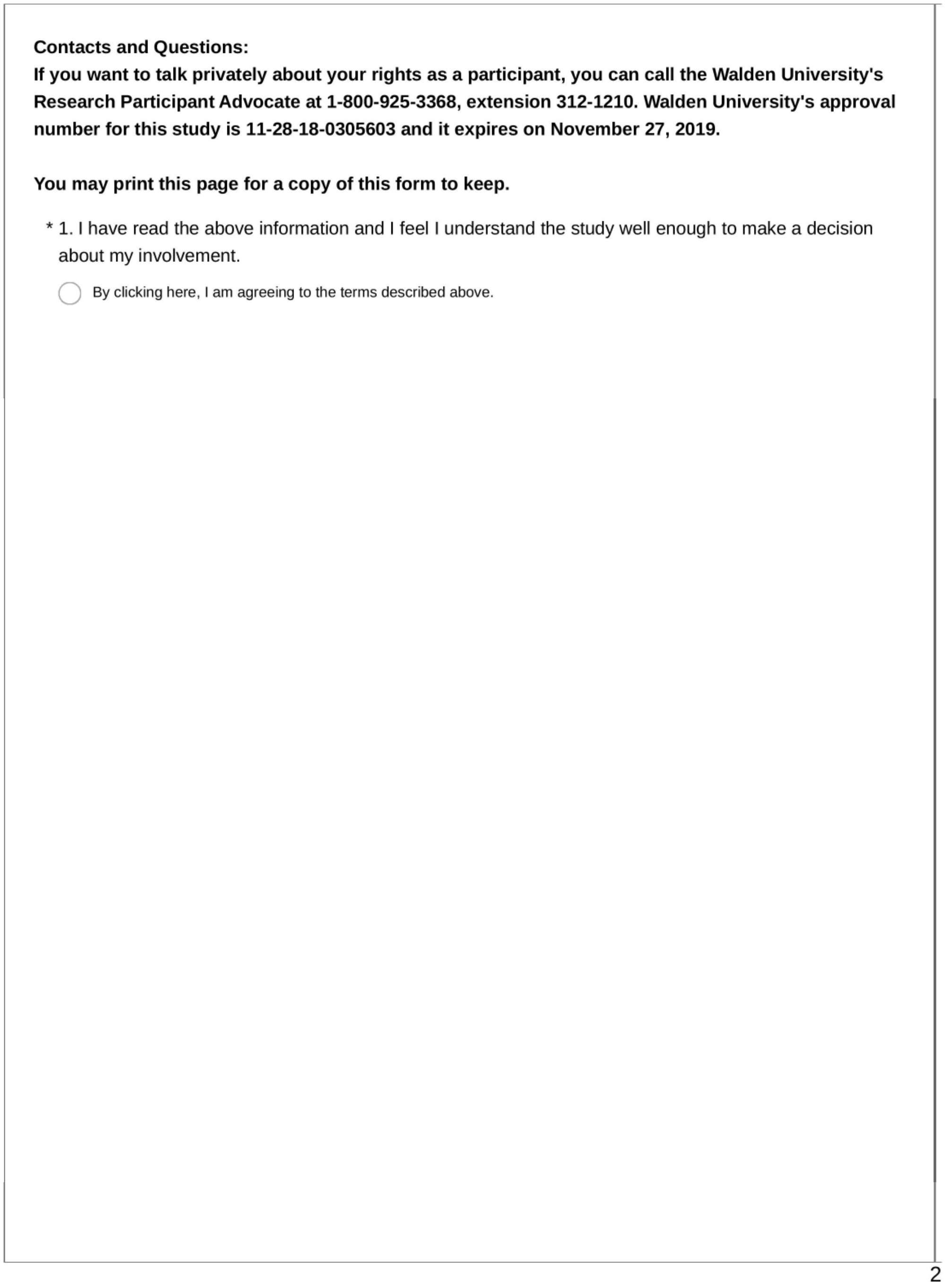
Procedures:
If you agree to participate in this study, after answering some questions of a general nature you will be asked to take an online survey by answering 39 questions. Participation in the online survey is expected to take 12 minutes.

Voluntary Nature of the Study:
Your participation in this study is voluntary. This means that everyone will respect your decision of whether or not you want to be in the study. If you decide to join the study now, you can still change your mind during the study. If you feel stressed during the study, you may stop at any time during the survey.

Risks and Benefits of Being in the Study:
The risks associated with participating in this study are minimal. The benefit of participating in the study is helping to gain further knowledge related to the research topic. You also will be given a link to download a free copy of the research results at the conclusion of this survey.

Compensation:
No compensation is being offered for participation in the study survey.

Confidentiality:
Any information you provide will be kept entirely anonymous. The researcher will not use your information for any purposes outside of this research project. Also, the researcher will not include your name or anything else that could identify you in any reports of the study. The study data will be maintained in a safe place for a minimum of five years to protect the rights of the participants. The data will be purged after the data retention period.

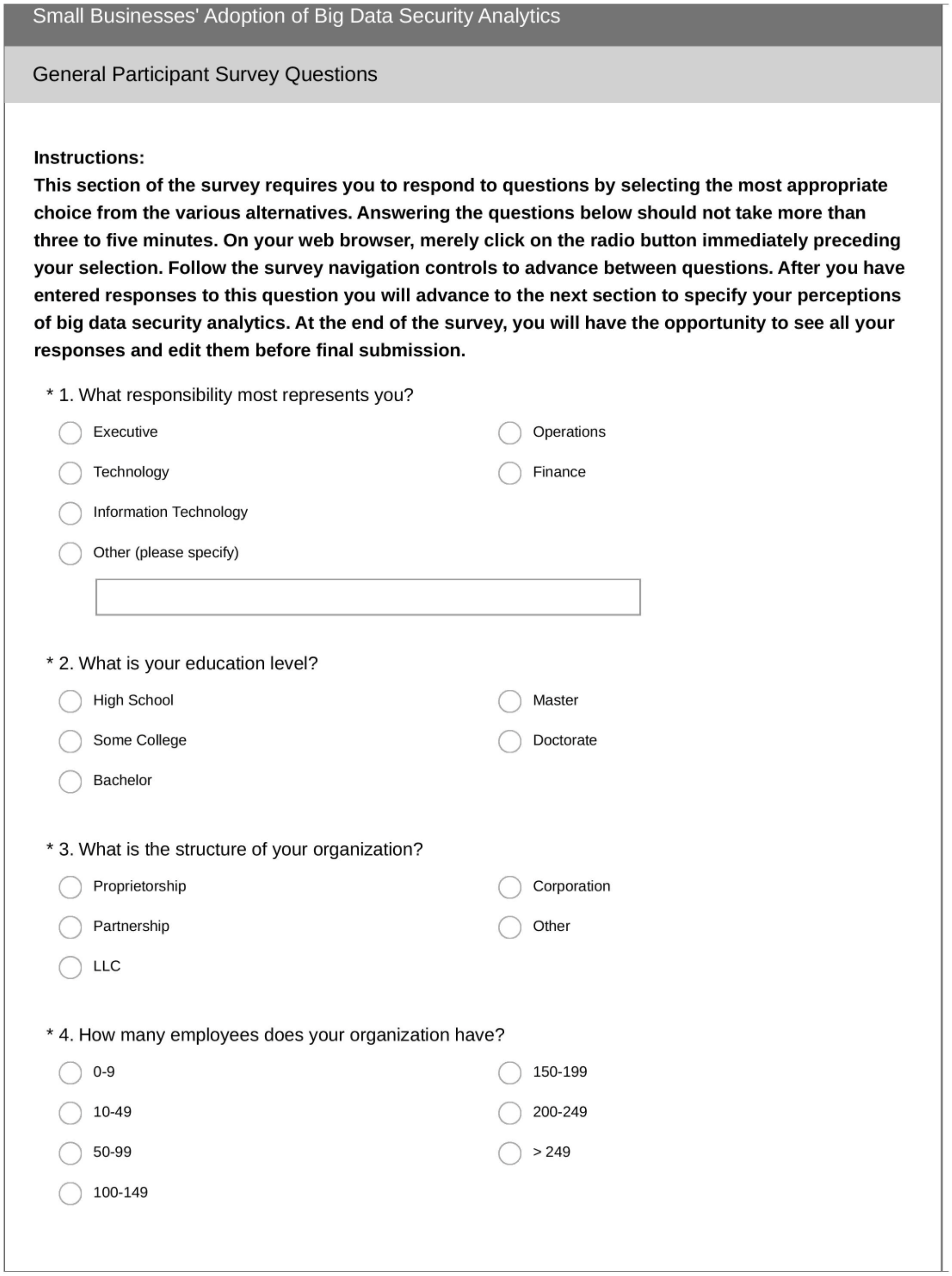


Contacts and Questions:
If you want to talk privately about your rights as a participant, you can call the Walden University's Research Participant Advocate at 1-800-925-3368, extension 312-1210. Walden University's approval number for this study is 11-28-18-0305603 and it expires on November 27, 2019.

You may print this page for a copy of this form to keep.

* 1. I have read the above information and I feel I understand the study well enough to make a decision about my involvement.

○ By clicking here, I am agreeing to the terms described above.



Small Businesses' Adoption of Big Data Security Analytics

General Participant Survey Questions

Instructions:
This section of the survey requires you to respond to questions by selecting the most appropriate choice from the various alternatives. Answering the questions below should not take more than three to five minutes. On your web browser, merely click on the radio button immediately preceding your selection. Follow the survey navigation controls to advance between questions. After you have entered responses to this question you will advance to the next section to specify your perceptions of big data security analytics. At the end of the survey, you will have the opportunity to see all your responses and edit them before final submission.

* 1. What responsibility most represents you?

○ Executive
○ Technology
○ Information Technology
○ Other (please specify)
○ Operations
○ Finance

* 2. What is your education level?

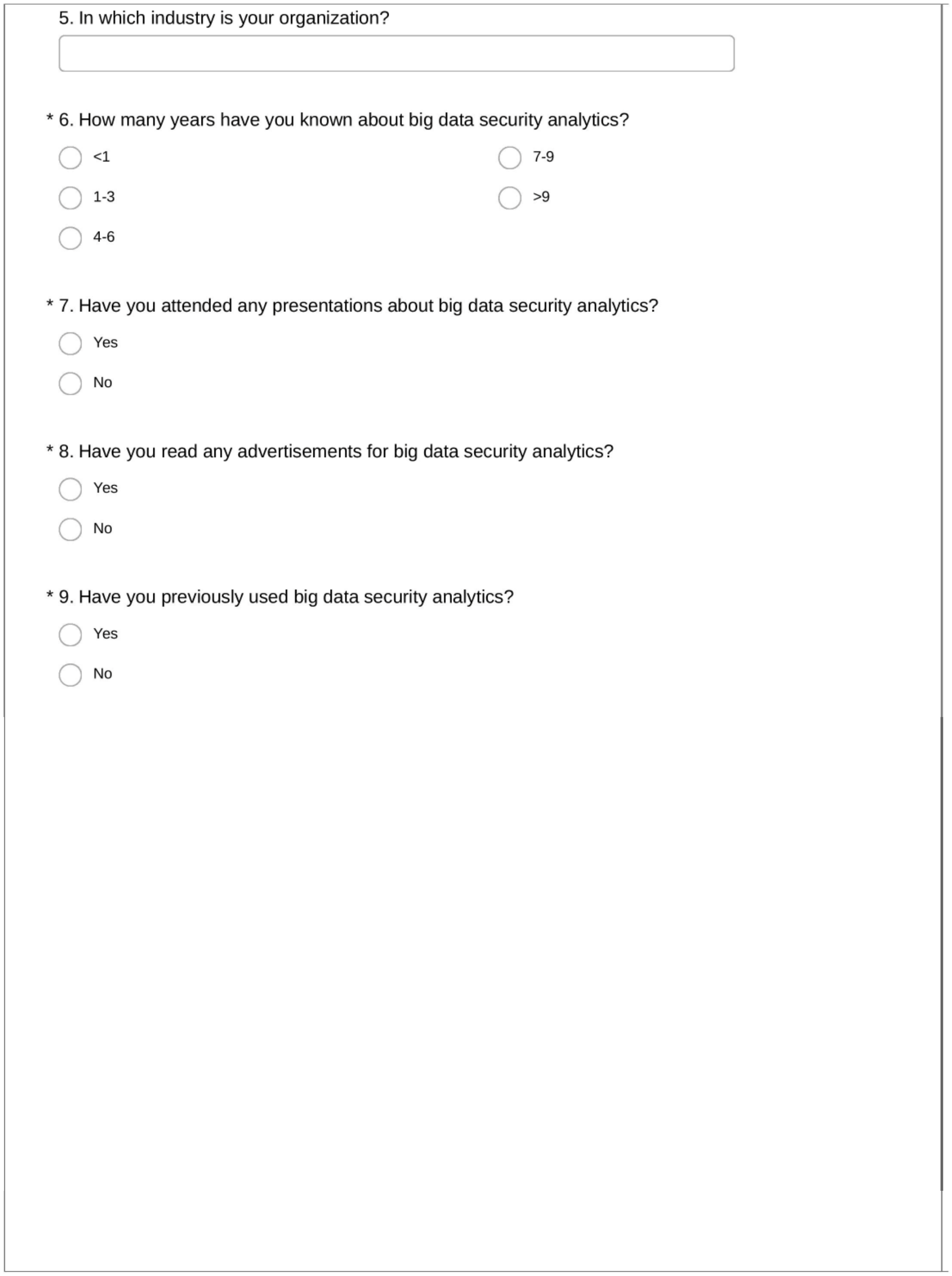
○ High School
○ Some College
○ Bachelor
○ Master
○ Doctorate

* 3. What is the structure of your organization?

○ Proprietorship
○ Partnership
○ LLC
○ Corporation
○ Other

* 4. How many employees does your organization have?

○ 0-9
○ 10-49
○ 50-99
○ 100-149
○ 150-199
○ 200-249
○ > 249



5. In which industry is your organization?

* 6. How many years have you known about big data security analytics?

○ <1

○ 1-3

○ 4-6

○ 7-9

○ >9

* 7. Have you attended any presentations about big data security analytics?

○ Yes

○ No

* 8. Have you read any advertisements for big data security analytics?

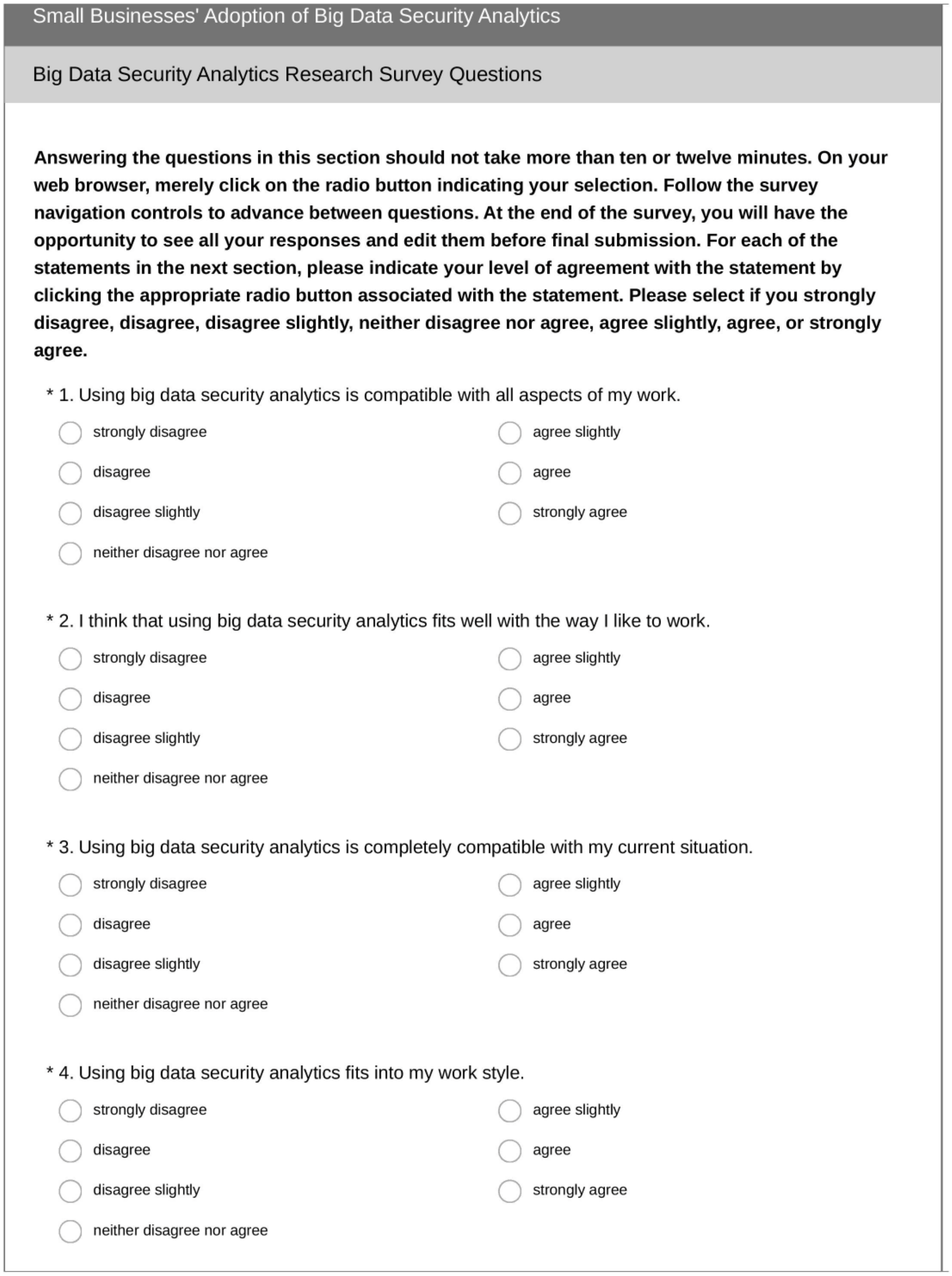
○ Yes

○ No

* 9. Have you previously used big data security analytics?

○ Yes

○ No



Small Businesses' Adoption of Big Data Security Analytics

Big Data Security Analytics Research Survey Questions

Answering the questions in this section should not take more than ten or twelve minutes. On your web browser, merely click on the radio button indicating your selection. Follow the survey navigation controls to advance between questions. At the end of the survey, you will have the opportunity to see all your responses and edit them before final submission. For each of the statements in the next section, please indicate your level of agreement with the statement by clicking the appropriate radio button associated with the statement. Please select if you strongly disagree, disagree, disagree slightly, neither disagree nor agree, agree slightly, agree, or strongly agree.

* 1. Using big data security analytics is compatible with all aspects of my work.

○ strongly disagree
○ disagree
○ disagree slightly
○ neither disagree nor agree
○ agree slightly
○ agree
○ strongly agree

* 2. I think that using big data security analytics fits well with the way I like to work.

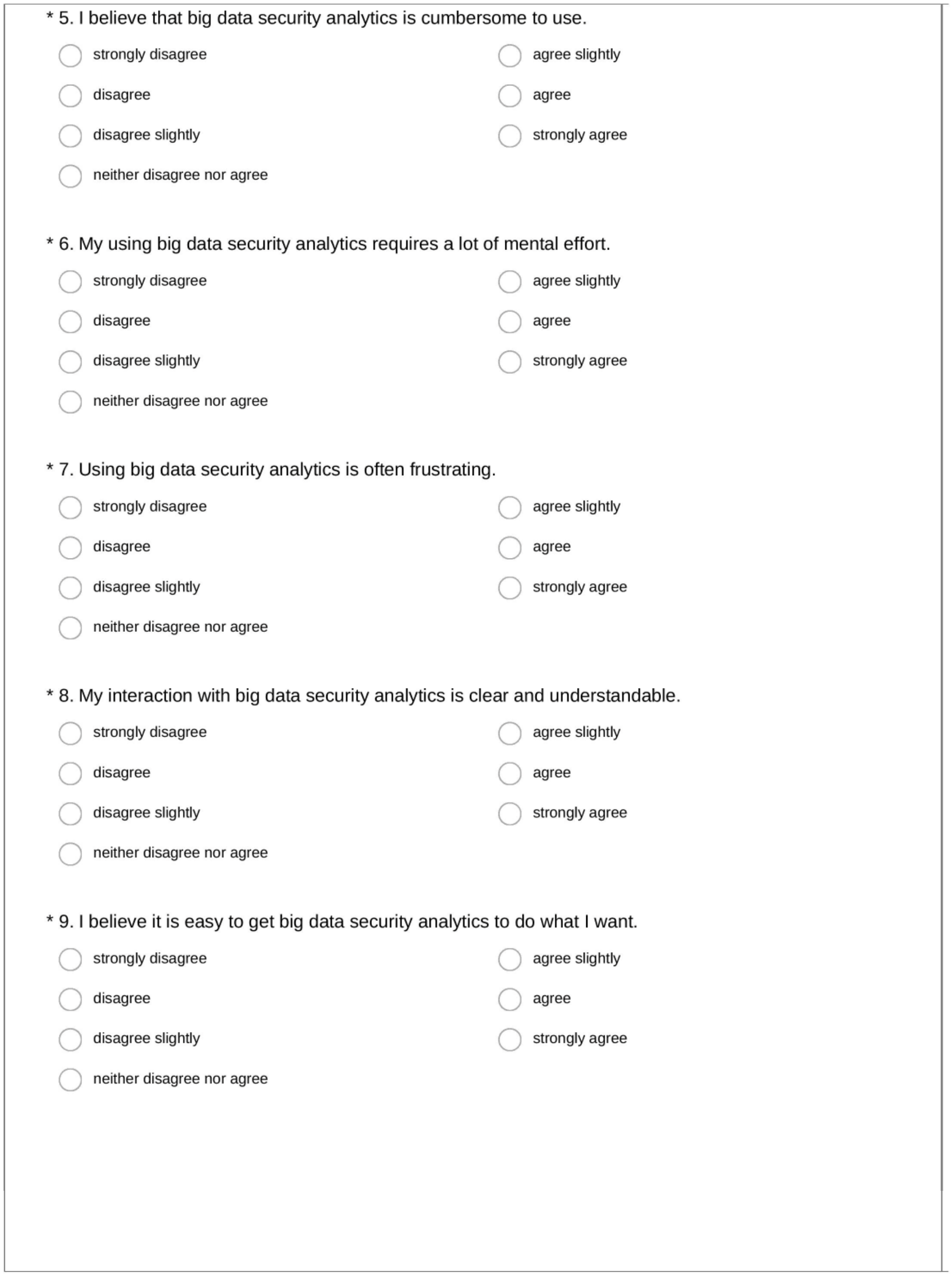
○ strongly disagree
○ disagree
○ disagree slightly
○ neither disagree nor agree
○ agree slightly
○ agree
○ strongly agree

* 3. Using big data security analytics is completely compatible with my current situation.

○ strongly disagree
○ disagree
○ disagree slightly
○ neither disagree nor agree
○ agree slightly
○ agree
○ strongly agree

* 4. Using big data security analytics fits into my work style.

○ strongly disagree
○ disagree
○ disagree slightly
○ neither disagree nor agree
○ agree slightly
○ agree
○ strongly agree



* 5. I believe that big data security analytics is cumbersome to use.

○ strongly disagree
○ disagree
○ disagree slightly
○ neither disagree nor agree
○ agree slightly
○ agree
○ strongly agree

* 6. My using big data security analytics requires a lot of mental effort.

○ strongly disagree
○ disagree
○ disagree slightly
○ neither disagree nor agree
○ agree slightly
○ agree
○ strongly agree

* 7. Using big data security analytics is often frustrating.

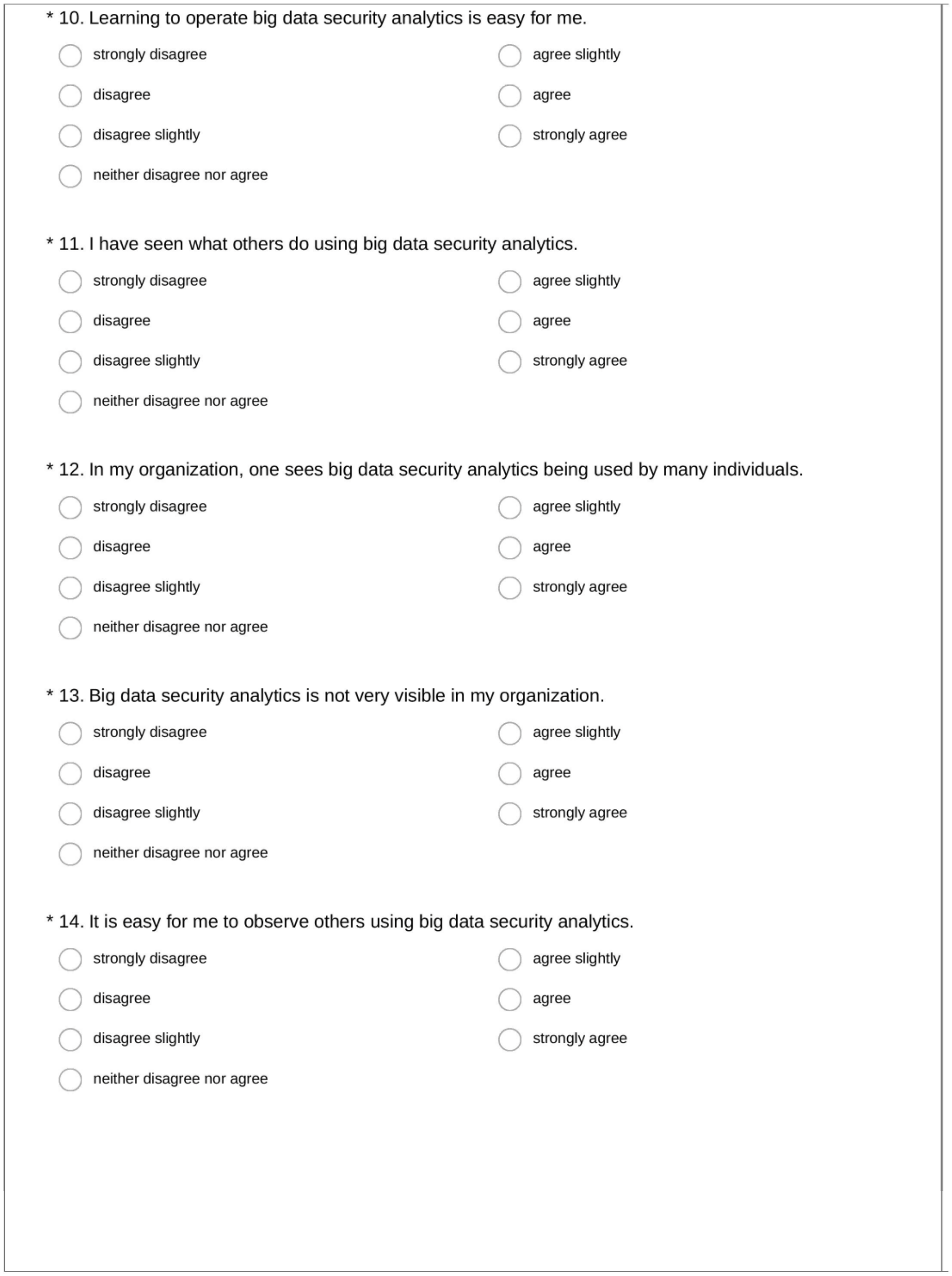
○ strongly disagree
○ disagree
○ disagree slightly
○ neither disagree nor agree
○ agree slightly
○ agree
○ strongly agree

* 8. My interaction with big data security analytics is clear and understandable.

○ strongly disagree
○ disagree
○ disagree slightly
○ neither disagree nor agree
○ agree slightly
○ agree
○ strongly agree

* 9. I believe it is easy to get big data security analytics to do what I want.

○ strongly disagree
○ disagree
○ disagree slightly
○ neither disagree nor agree
○ agree slightly
○ agree
○ strongly agree



* 10. Learning to operate big data security analytics is easy for me.

○ strongly disagree
○ disagree
○ disagree slightly
○ neither disagree nor agree
○ agree slightly
○ agree
○ strongly agree

* 11. I have seen what others do using big data security analytics.

○ strongly disagree
○ disagree
○ disagree slightly
○ neither disagree nor agree
○ agree slightly
○ agree
○ strongly agree

* 12. In my organization, one sees big data security analytics being used by many individuals.

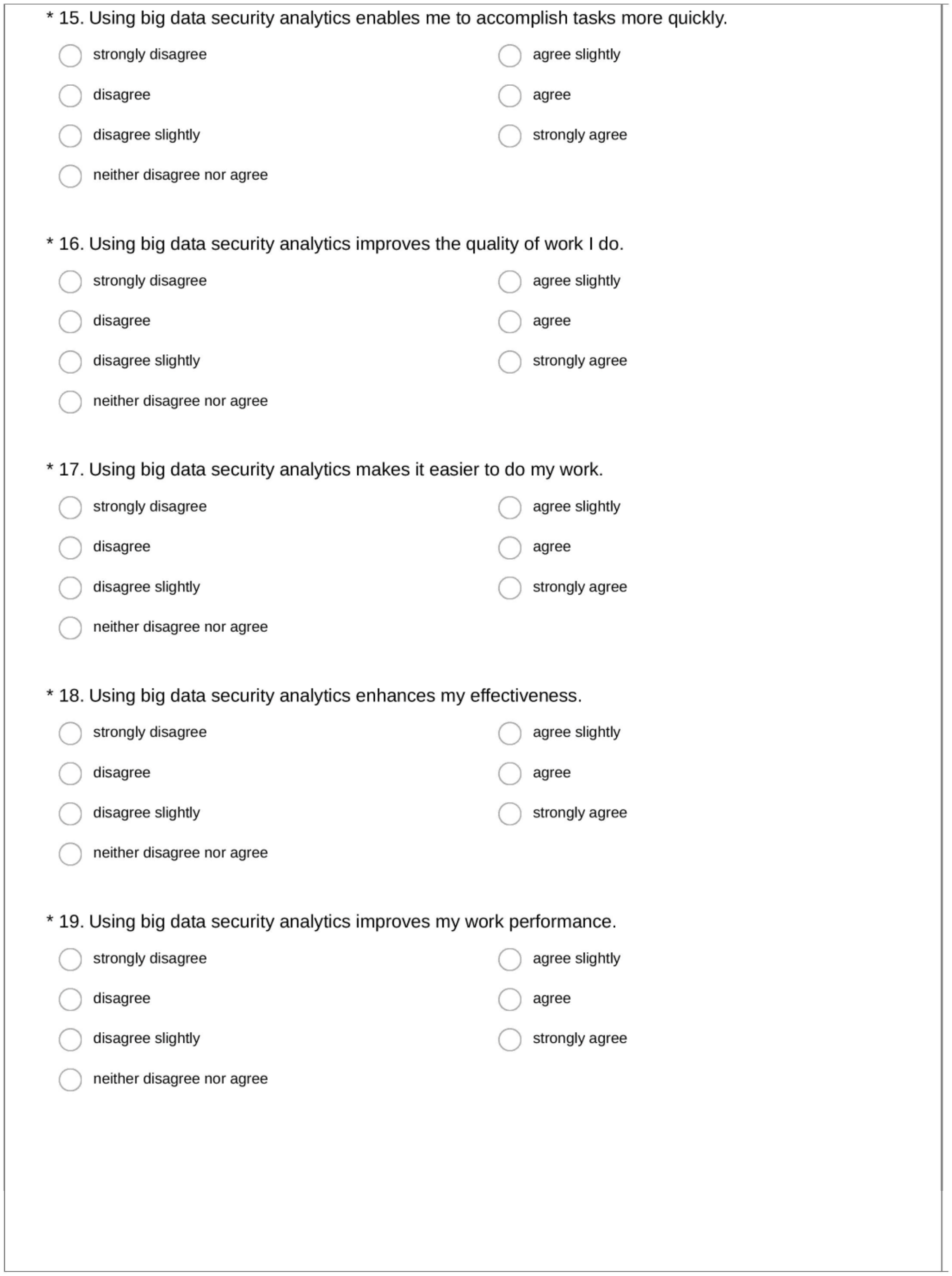
○ strongly disagree
○ disagree
○ disagree slightly
○ neither disagree nor agree
○ agree slightly
○ agree
○ strongly agree

* 13. Big data security analytics is not very visible in my organization.

○ strongly disagree
○ disagree
○ disagree slightly
○ neither disagree nor agree
○ agree slightly
○ agree
○ strongly agree

* 14. It is easy for me to observe others using big data security analytics.

○ strongly disagree
○ disagree
○ disagree slightly
○ neither disagree nor agree
○ agree slightly
○ agree
○ strongly agree



* 15. Using big data security analytics enables me to accomplish tasks more quickly.

○ strongly disagree
○ disagree
○ disagree slightly
○ neither disagree nor agree
○ agree slightly
○ agree
○ strongly agree

* 16. Using big data security analytics improves the quality of work I do.

○ strongly disagree
○ disagree
○ disagree slightly
○ neither disagree nor agree
○ agree slightly
○ agree
○ strongly agree

* 17. Using big data security analytics makes it easier to do my work.

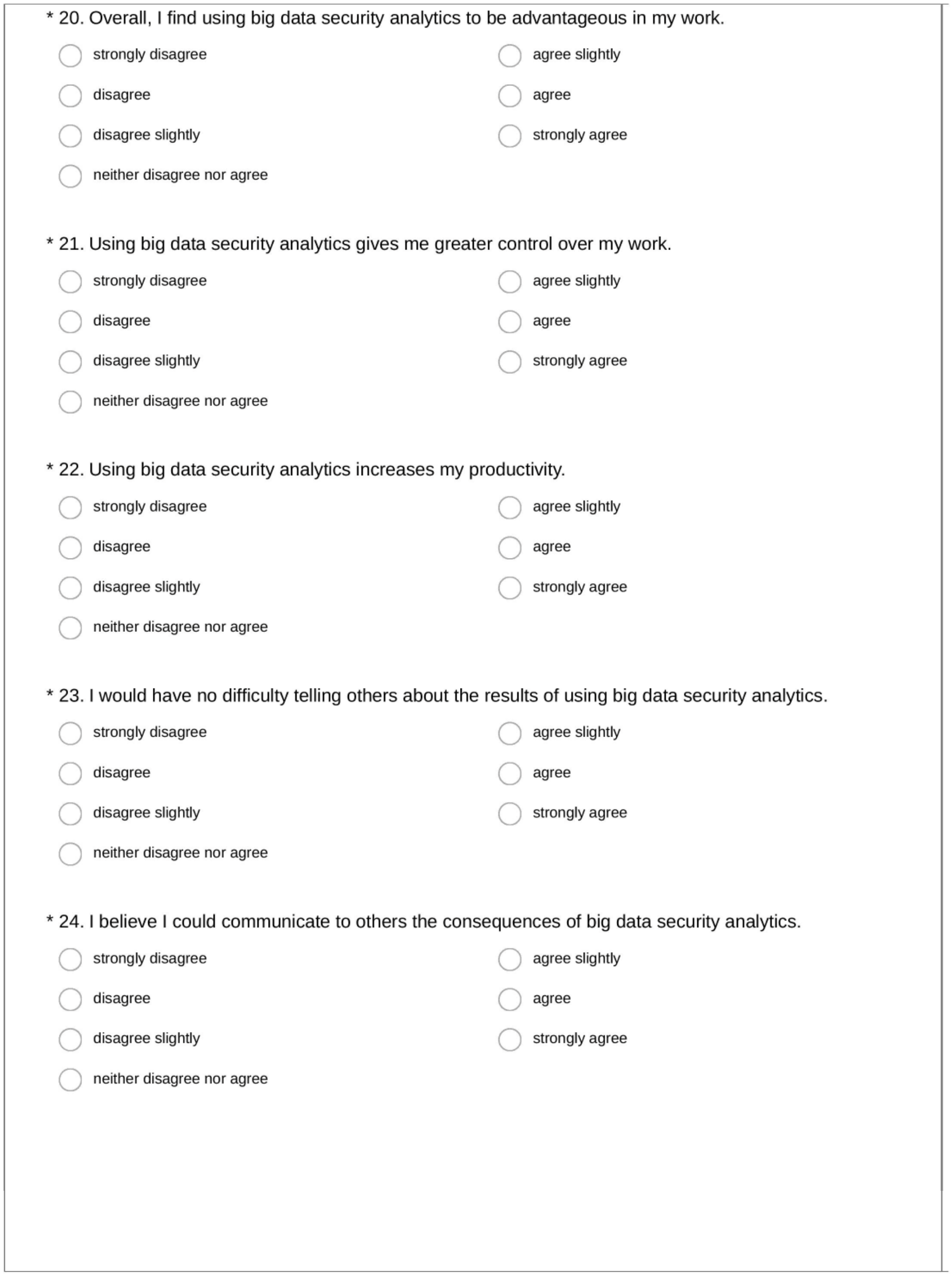
○ strongly disagree
○ disagree
○ disagree slightly
○ neither disagree nor agree
○ agree slightly
○ agree
○ strongly agree

* 18. Using big data security analytics enhances my effectiveness.

○ strongly disagree
○ disagree
○ disagree slightly
○ neither disagree nor agree
○ agree slightly
○ agree
○ strongly agree

* 19. Using big data security analytics improves my work performance.

○ strongly disagree
○ disagree
○ disagree slightly
○ neither disagree nor agree
○ agree slightly
○ agree
○ strongly agree



* 20. Overall, I find using big data security analytics to be advantageous in my work.

○ strongly disagree
○ disagree
○ disagree slightly
○ neither disagree nor agree
○ agree slightly
○ agree
○ strongly agree

* 21. Using big data security analytics gives me greater control over my work.

○ strongly disagree
○ disagree
○ disagree slightly
○ neither disagree nor agree
○ agree slightly
○ agree
○ strongly agree

* 22. Using big data security analytics increases my productivity.

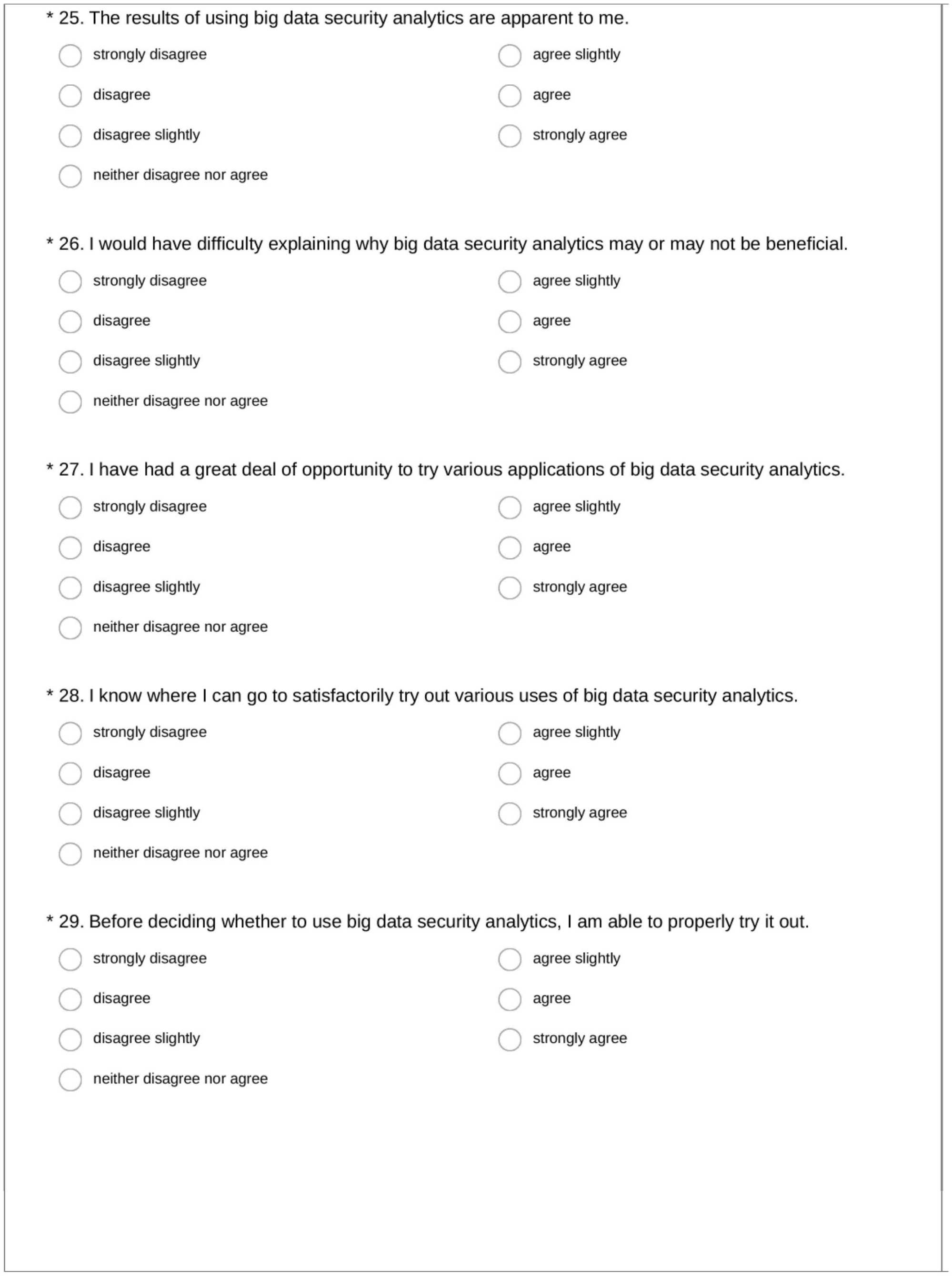
○ strongly disagree
○ disagree
○ disagree slightly
○ neither disagree nor agree
○ agree slightly
○ agree
○ strongly agree

* 23. I would have no difficulty telling others about the results of using big data security analytics.

○ strongly disagree
○ disagree
○ disagree slightly
○ neither disagree nor agree
○ agree slightly
○ agree
○ strongly agree

* 24. I believe I could communicate to others the consequences of big data security analytics.

○ strongly disagree
○ disagree
○ disagree slightly
○ neither disagree nor agree
○ agree slightly
○ agree
○ strongly agree



* 25. The results of using big data security analytics are apparent to me.

○ strongly disagree
○ disagree
○ disagree slightly
○ neither disagree nor agree
○ agree slightly
○ agree
○ strongly agree

* 26. I would have difficulty explaining why big data security analytics may or may not be beneficial.

○ strongly disagree
○ disagree
○ disagree slightly
○ neither disagree nor agree
○ agree slightly
○ agree
○ strongly agree

* 27. I have had a great deal of opportunity to try various applications of big data security analytics.

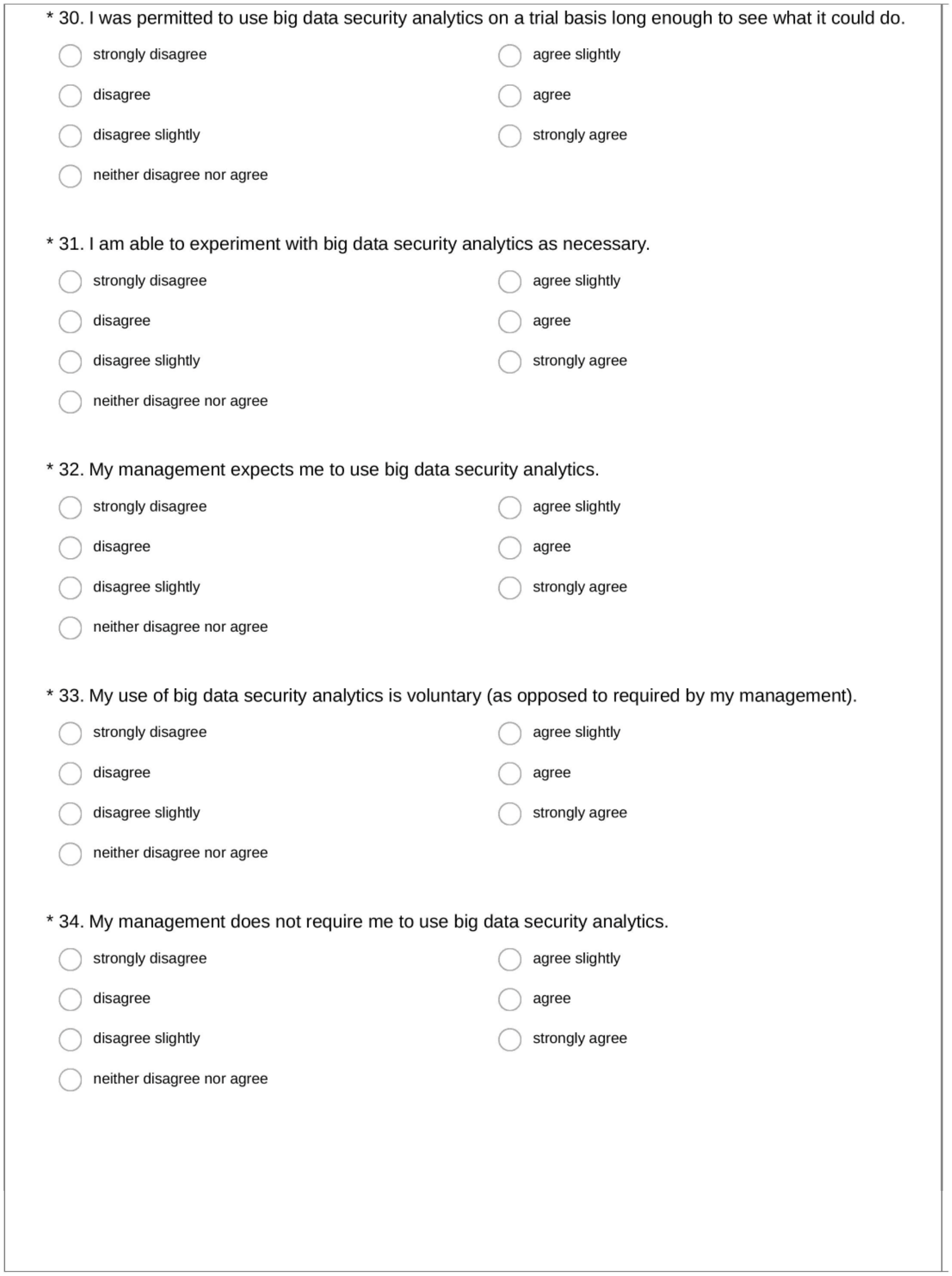
○ strongly disagree
○ disagree
○ disagree slightly
○ neither disagree nor agree
○ agree slightly
○ agree
○ strongly agree

* 28. I know where I can go to satisfactorily try out various uses of big data security analytics.

○ strongly disagree
○ disagree
○ disagree slightly
○ neither disagree nor agree
○ agree slightly
○ agree
○ strongly agree

* 29. Before deciding whether to use big data security analytics, I am able to properly try it out.

○ strongly disagree
○ disagree
○ disagree slightly
○ neither disagree nor agree
○ agree slightly
○ agree
○ strongly agree



* 30. I was permitted to use big data security analytics on a trial basis long enough to see what it could do.

○ strongly disagree
○ disagree
○ disagree slightly
○ neither disagree nor agree
○ agree slightly
○ agree
○ strongly agree

* 31. I am able to experiment with big data security analytics as necessary.

○ strongly disagree
○ disagree
○ disagree slightly
○ neither disagree nor agree
○ agree slightly
○ agree
○ strongly agree

* 32. My management expects me to use big data security analytics.

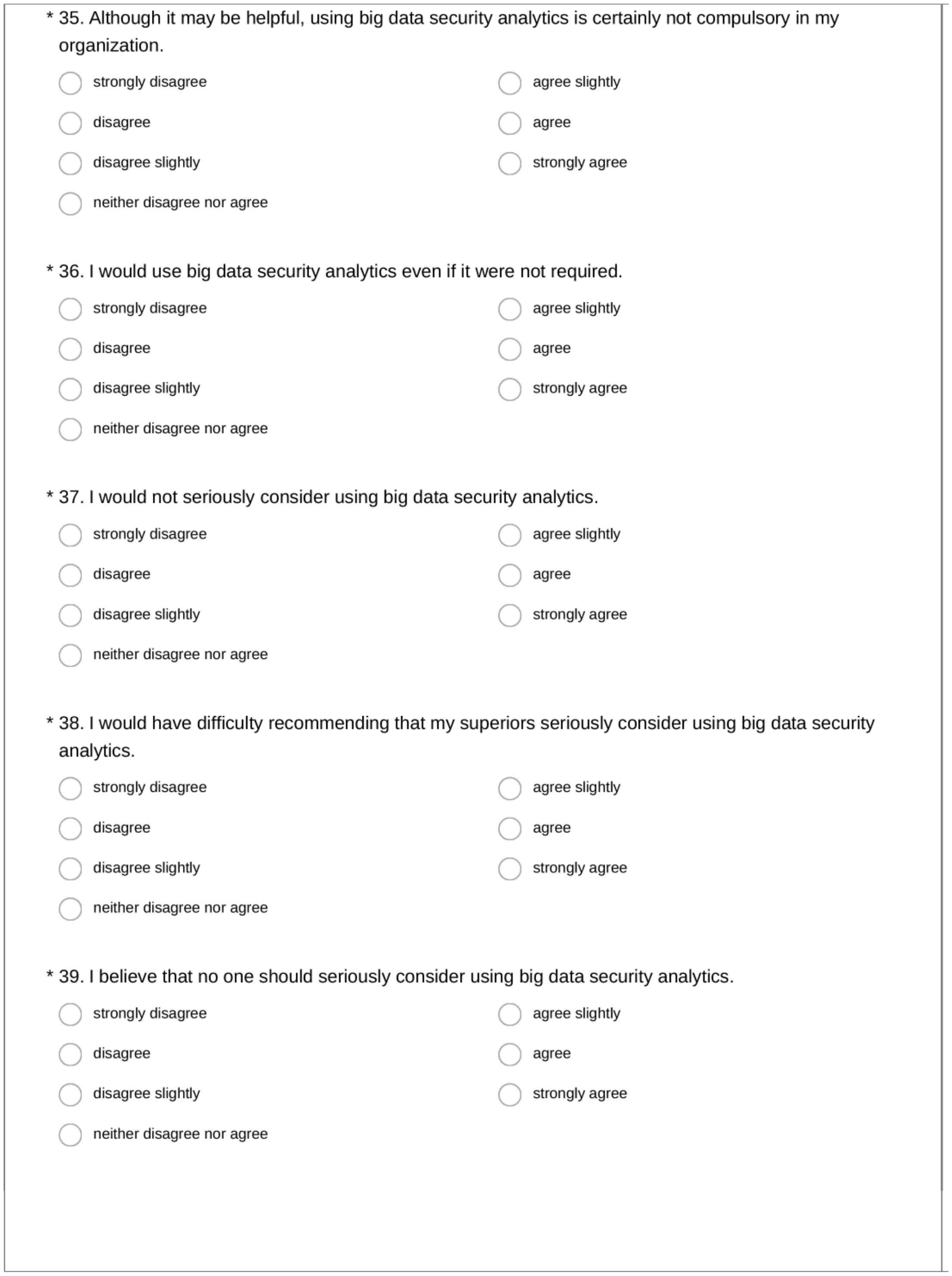
○ strongly disagree
○ disagree
○ disagree slightly
○ neither disagree nor agree
○ agree slightly
○ agree
○ strongly agree

* 33. My use of big data security analytics is voluntary (as opposed to required by my management).

○ strongly disagree
○ disagree
○ disagree slightly
○ neither disagree nor agree
○ agree slightly
○ agree
○ strongly agree

* 34. My management does not require me to use big data security analytics.

○ strongly disagree
○ disagree
○ disagree slightly
○ neither disagree nor agree
○ agree slightly
○ agree
○ strongly agree



* 35. Although it may be helpful, using big data security analytics is certainly not compulsory in my organization.

○ strongly disagree
○ disagree
○ disagree slightly
○ neither disagree nor agree
○ agree slightly
○ agree
○ strongly agree

* 36. I would use big data security analytics even if it were not required.

○ strongly disagree
○ disagree
○ disagree slightly
○ neither disagree nor agree
○ agree slightly
○ agree
○ strongly agree

* 37. I would not seriously consider using big data security analytics.

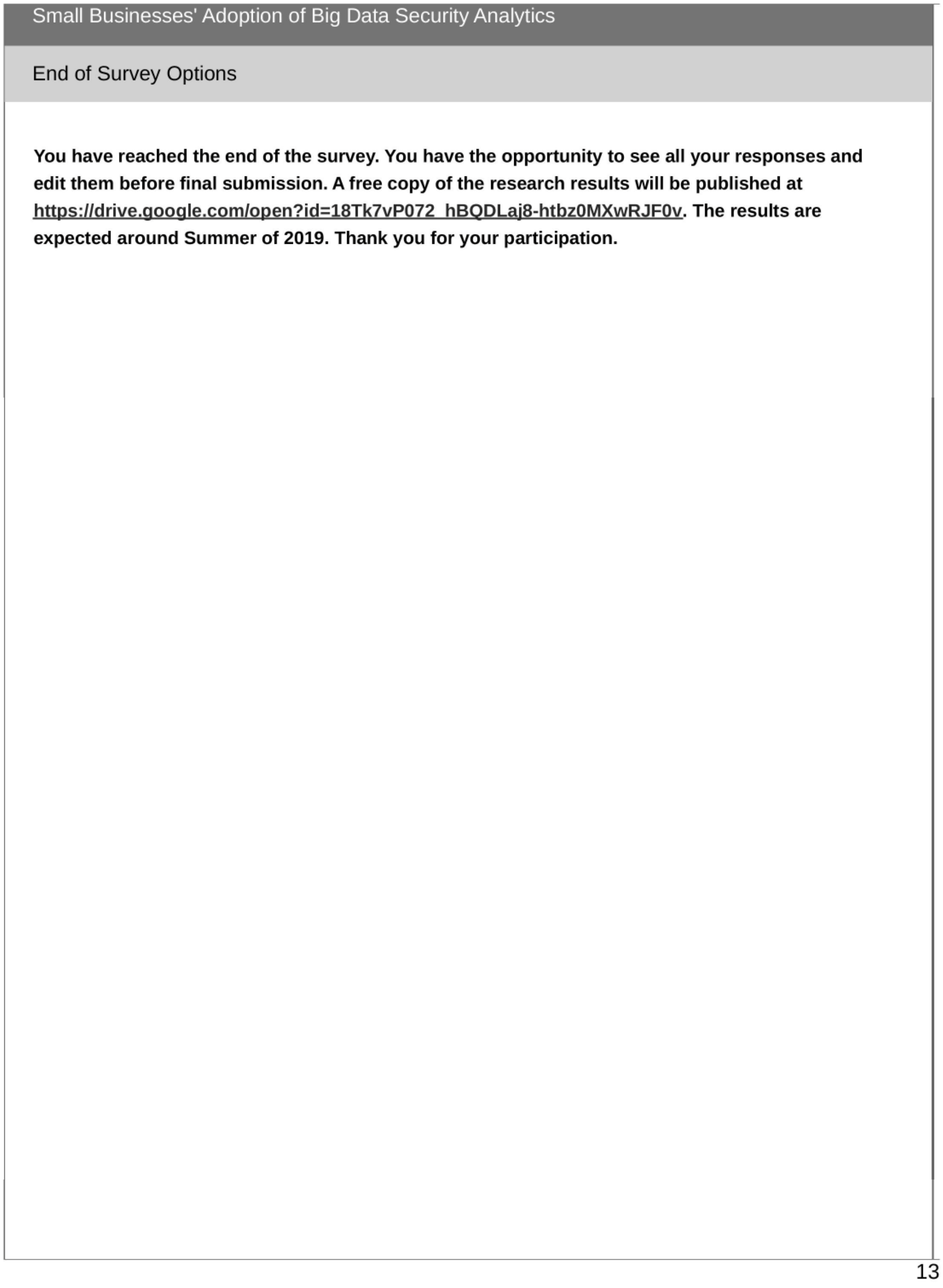
○ strongly disagree
○ disagree
○ disagree slightly
○ neither disagree nor agree
○ agree slightly
○ agree
○ strongly agree

* 38. I would have difficulty recommending that my superiors seriously consider using big data security analytics.

○ strongly disagree
○ disagree
○ disagree slightly
○ neither disagree nor agree
○ agree slightly
○ agree
○ strongly agree

* 39. I believe that no one should seriously consider using big data security analytics.

○ strongly disagree
○ disagree
○ disagree slightly
○ neither disagree nor agree
○ agree slightly
○ agree
○ strongly agree



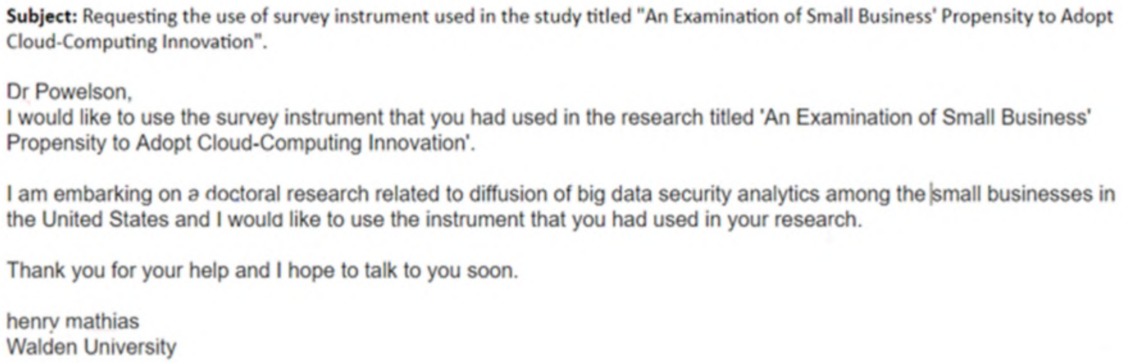
Small Businesses' Adoption of Big Data Security Analytics

End of Survey Options

You have reached the end of the survey. You have the opportunity to see all your responses and edit them before final submission. A free copy of the research results will be published at https://drive.google.com/open?id=18Tk7vP072_hBQDLaj8-htbz0MXwRJF0v. The results are expected around Summer of 2019. Thank you for your participation.

13

Appendix B: Request for Approval to Use Survey Instrument



Subject: Requesting the use of survey instrument used in the study titled "An Examination of Small Business' Propensity to Adopt Cloud-Computing Innovation".

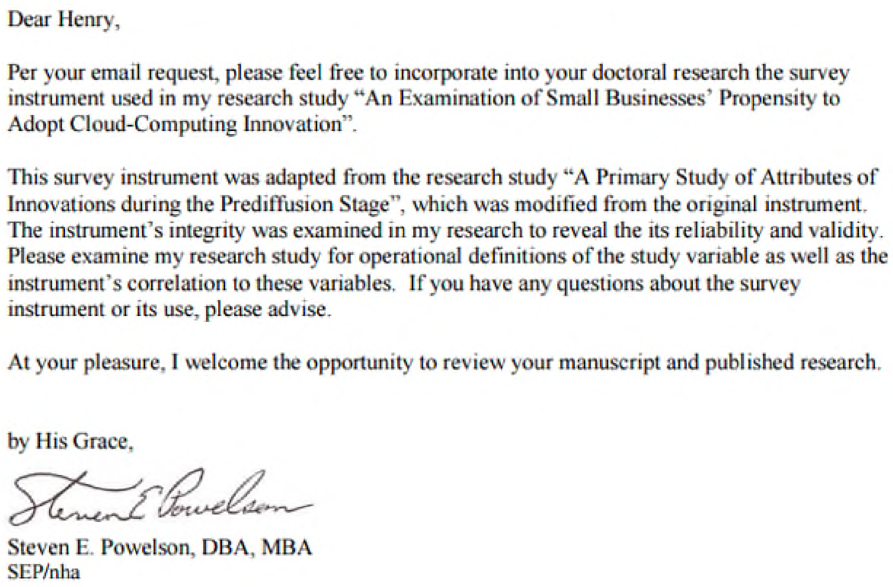
Dr Powelson,
I would like to use the survey instrument that you had used in the research titled 'An Examination of Small Business' Propensity to Adopt Cloud-Computing Innovation'.

I am embarking on a doctoral research related to diffusion of big data security analytics among the small businesses in the United States and I would like to use the instrument that you had used in your research.

Thank you for your help and I hope to talk to you soon.

henry mathias
Walden University

Appendix C: Approval to Use Survey Instrument



Dear Henry,

Per your email request, please feel free to incorporate into your doctoral research the survey instrument used in my research study “An Examination of Small Businesses’ Propensity to Adopt Cloud-Computing Innovation”.

This survey instrument was adapted from the research study “A Primary Study of Attributes of Innovations during the Prediffusion Stage”, which was modified from the original instrument. The instrument’s integrity was examined in my research to reveal the its reliability and validity. Please examine my research study for operational definitions of the study variable as well as the instrument’s correlation to these variables. If you have any questions about the survey instrument or its use, please advise.

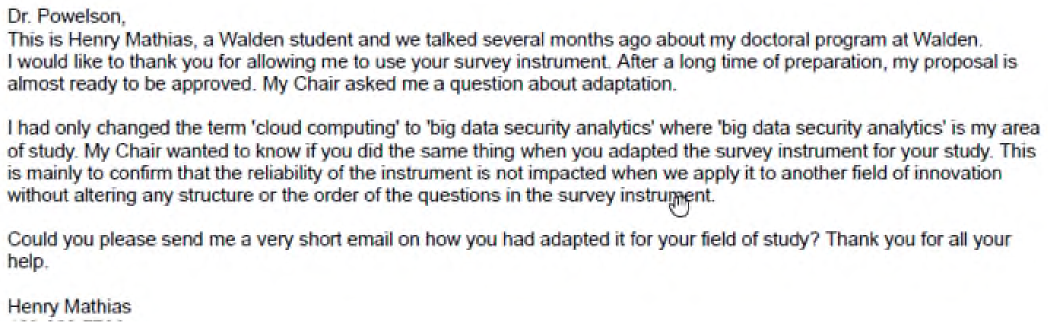
At your pleasure, I welcome the opportunity to review your manuscript and published research.

by His Grace,

Steven E Powelson

Steven E. Powelson, DBA, MBA
SEP/nha

Appendix D: Request for the Method of Adaptation of Survey Instrument



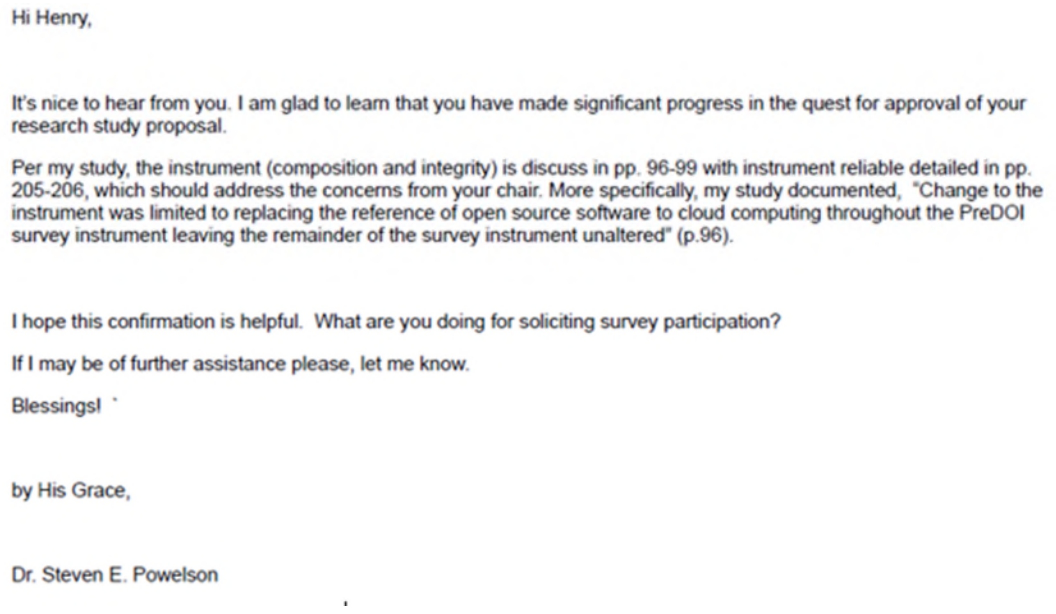
Dr. Powelson,
This is Henry Mathias, a Walden student and we talked several months ago about my doctoral program at Walden.
I would like to thank you for allowing me to use your survey instrument. After a long time of preparation, my proposal is almost ready to be approved. My Chair asked me a question about adaptation.

I had only changed the term 'cloud computing' to 'big data security analytics' where 'big data security analytics' is my area of study. My Chair wanted to know if you did the same thing when you adapted the survey instrument for your study. This is mainly to confirm that the reliability of the instrument is not impacted when we apply it to another field of innovation without altering any structure or the order of the questions in the survey instrument.

Could you please send me a very short email on how you had adapted it for your field of study? Thank you for all your help.

Henry Mathias

Appendix E: Response for the Method of Adaptation of Survey Instrument



Hi Henry,

It's nice to hear from you. I am glad to learn that you have made significant progress in the quest for approval of your research study proposal.

Per my study, the instrument (composition and integrity) is discuss in pp. 96-99 with instrument reliable detailed in pp. 205-206, which should address the concerns from your chair. More specifically, my study documented, "Change to the instrument was limited to replacing the reference of open source software to cloud computing throughout the PreDOI survey instrument leaving the remainder of the survey instrument unaltered" (p.96).

I hope this confirmation is helpful. What are you doing for soliciting survey participation?

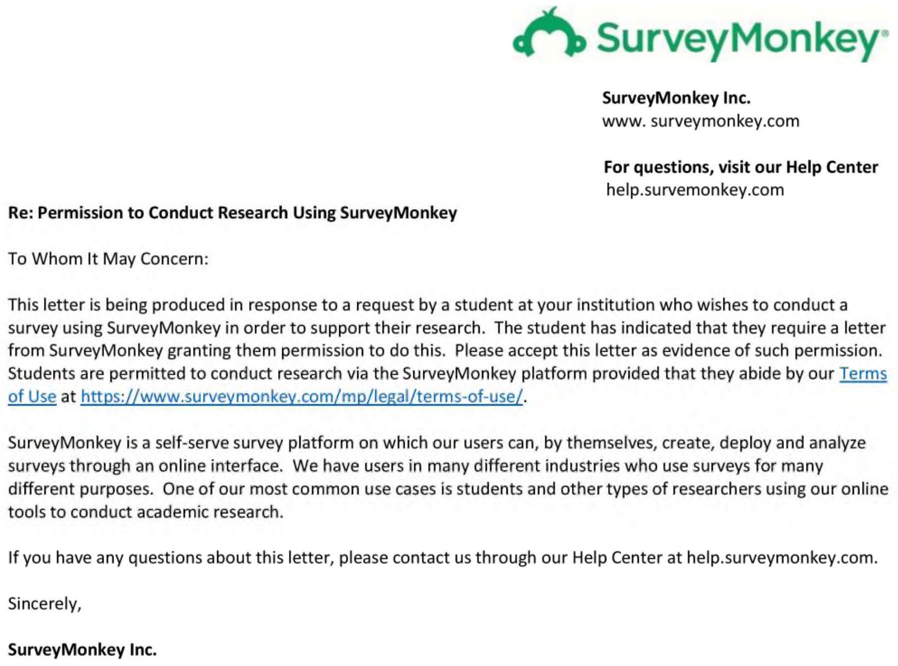
If I may be of further assistance please, let me know.

Blessings!

by His Grace,

Dr. Steven E. Powelson

Appendix F: Survey Monkey Permission



SurveyMonkey Inc.
www. surveymonkey.com

For questions, visit our Help Center
help.survemonkey.com

Re: Permission to Conduct Research Using SurveyMonkey

To Whom It May Concern:

This letter is being produced in response to a request by a student at your institution who wishes to conduct a survey using SurveyMonkey in order to support their research. The student has indicated that they require a letter from SurveyMonkey granting them permission to do this. Please accept this letter as evidence of such permission. Students are permitted to conduct research via the SurveyMonkey platform provided that they abide by our Terms of Use at https://www.surveymonkey.com/mp/legal/terms-of-use/.

SurveyMonkey is a self-serve survey platform on which our users can, by themselves, create, deploy and analyze surveys through an online interface. We have users in many different industries who use surveys for many different purposes. One of our most common use cases is students and other types of researchers using our online tools to conduct academic research.

If you have any questions about this letter, please contact us through our Help Center at help.surveymonkey.com.

Sincerely,

SurveyMonkey Inc.

Appendix G: General Respondent Information Descriptive Statistics

Table G1

Participant Characteristic Descriptive Statistics by Responsibility

			Valid Percent	
Responsibility	Frequency	Percent	Item	Cumulative
Other	1	.6	.6	.6
Executive	53	32.1	32.1	32.7
Technology	33	20.0	20.0	52.7
Information technology	61	37.0	37.0	89.7
Operations	10	6.1	6.1	95.8
Finance	7	4.2	4.2	100.0
Total	165	100.0	100.0	

Note. No missing values.

Table G2

Participant Characteristic Descriptive Statistics by Age

			Valid Percent	
Age range	Frequency	Percent	Item	Cumulative
18-29	28	17.0	17.0	17.0
30-44	89	53.9	53.9	70.9
45-60	29	17.6	17.6	88.5
60+	19	11.5	11.5	100.0
Total	165	100.0	100.0	

Note. No missing values.

Table G3

Participant Characteristic Descriptive Statistics by Education

			Valid Percent	
Gender	Frequency	Percent	Item	Cumulative
High school	12	7.3	7.3	7.3
Some college	25	15.2	15.2	22.4
Bachelor	70	42.4	42.4	64.8
Master	47	28.5	28.5	93.3
Doctorate	11	6.7	6.7	100.0
Total	165	100.0	100.0	

Note. No missing values.

Appendix H: General Small Business Descriptive Statistics

Table H1

Small Business Attributes Descriptive Statistics by Category

			Valid Percent	
Legal structure	Frequency	Percent	Item	Cumulative
Proprietorship	37	22.4	22.4	22.4
Partnership	26	15.8	15.8	38.2
LLC	32	19.4	19.4	57.6
Corporation	63	38.2	38.2	95.8
Other	7	4.2	4.2	100.0
Total	165	100.0	100.0	

Note. No missing values.

Table H2

Small Business Attributes Descriptive Statistics by Number of Employees

			Valid Percent	
Number employees	Frequency	Percent	Item	Cumulative
0-9	9	5.5	5.5	5.5
10-49	36	21.8	21.8	27.3
50-99	39	23.6	23.6	50.9
100-149	39	23.6	23.6	74.5
150-199	28	17.0	17.0	91.5
200-249	14	8.5	8.5	100.0
Total	165	100.0	100.0	

Note. No missing values.

Table H3

Small Business Attributes Descriptive Statistics by Industry

			Valid Percent	
Industry	Frequency	Percent	Item	Cumulative
Agri., Forest, Fishing & Hunting	2	1.2	1.2	1.2
Mining, Quarrying, & Oil & Gas	1	.6	.6	1.8
Utilities	4	2.4	2.4	4.2
Construction	26	15.8	15.8	20.0
Manufacturing	8	4.8	4.8	24.8
Wholesale Trade	3	1.8	1.8	26.7
Retail Trade	8	4.8	4.8	31.5
Transport & Warehousing	3	1.8	1.8	33.3
Information	28	17.0	17.0	50.3
Finance & Insurance	10	6.1	6.1	56.4
Real restate & Rent, & Lease	3	1.8	1.8	58.2
Professional, Sci., & Tech., Services	34	20.6	20.6	78.8
Management of Companies	6	3.6	3.6	82.4
Admin., & Waste, & Remediation	6	3.6	3.6	86.1
Educational Services	4	2.4	2.4	88.5
Health., and social assistance	7	4.2	4.2	92.7
Arts, Entertainment, & Recreation	2	1.2	1.2	93.9
Other Services	7	4.2	4.2	98.2
Public Administration	3	1.8	1.8	100.0
Total	165	100.0	100.0	

Note. No missing values.

Appendix I: General Small Business Descriptive Statistics

Table I1

Big Data Security Analytics Awareness Descriptive Statistics

Extent of Awareness			Valid Percent	
Category	Frequency	Percent	Item	Cumulative
6. Years known about big data security analytics				
<1	7	4.2	4.2	4.2
1-3	36	21.8	21.8	26.1
4-6	72	43.6	43.6	69.7
7-9	32	19.4	19.4	89.1
>9	18	10.9	10.9	100.0
Total	165	100.0	100.0	
7. Attended big data security analytics’ presentation				
Yes	120	72.7	72.7	72.7
No	45	27.3	27.3	100.0
Total	165	100.0	100.0	
8. Read big data security analytics’ advertisement				
Yes	131	79.4	79.4	79.4
No	34	20.6	20.6	100.0
Total	165			
9. Previously used big data security analytics				
Yes	127	77.0	77.0	77.0
No	38	23.0	23.0	100.0
Total	165	100.0	100.0	

Note. No missing values.

Appendix J: Study Variable Descriptive Statistics

Table J1

Study Questions Descriptive Within Study Variable

	Range				
Variable Question [a]	Min.	Max.	Mean	Std. deviation	Variance
Compatibility					
Q1	1	7	5.430	1.453	2.112
Q2	2	7	5.781	1.099	1.208
Q3	2	7	5.769	1.102	1.215
Q4	3	7	5.836	1.055	1.113
Complexity					
Q5	1	7	4.842	1.721	2.963
Q6	1	7	5.097	1.419	2.015
Q7	1	7	4.339	1.740	3.030
Q8	1	7	5.587	1.131	1.280
Q9	2	7	5.539	1.176	1.384
Q10	2	7	5.454	1.139	1.298
Observability					
Q11	2	7	5.684	1.028	1.059
Q12	2	7	5.297	1.307	1.710
Q13	1	7	4.527	1.875	3.519
Q14	2	7	5.424	1.235	1.526
Relative Advantage					
Q15	2	7	5.642	1.136	1.262
Q16	2	7	5.715	1.103	1.217
Q17	1	7	5.781	1.082	1.172
Q18	2	7	5.787	1.016	1.034
Q19	1	7	5.690	1.207	1.459
Q20	2	7	5.793	1.067	1.140
Q21	1	7	5.697	1.206	1.456
Q22	1	7	5.793	1.155	1.335
Trialability					
Q27	2	7	5.224	1.354	1.834
Q28	2	7	5.497	1.156	1.337
Q29	3	7	5.648	1.092	1.193
Q30	1	7	5.515	1.276	1.629
Q31	2	7	5.757	1.143	1.307

Note. [a] n = 165.

(*table continues*)

	Range				
Variable Question [a]	Min.	Max.	Mean	Std. deviation	Variance
Adoption					
Q36	1	7	5.472	1.295	1.678
Q37	1	7	4.018	2.142	4.591
Q38	1	7	4.187	1.939	3.763
Q39	1	7	4.084	2.231	4.981

Note. [a] n = 165.

Table J2

Study Variable Central Tendency and Dispersion Statistics Within Type

	Range				
Variable Question [a]	Min.	Max.	Mean	Std. deviation	Variance
Predictors					
Relative	3.38	7.00	5.737	.908	.825
Advantage					
Compatibility	3.00	7.00	5.704	.954	.910
Complexity	2.00	4.83	3.616	.652	.425
Observability	3.25	7.00	4.969	.803	.645
Trialability	3.00	7.00	5.528	.937	.879
Criterion					
Adoption	1.00	7.00	4.295	1.457	2.123

Note. [a] n = 165.

Table J3

Study Variable Distribution Statistics within Type

Type Question [a]	Skewness [b]	Kurtosis [c]
Predictor variables		
Relative Advantage	-.720	-.295
Compatibility	-.693	-.194
Complexity	-.690	-.184
Observability	.263	-.054
Trialability	-.533	-.516
Criterion variable		
Adoption	.354	-1.076

Note. [a] n = 165. [b] Standard error = .189. [c] Standard error = .376.

Appendix K: Linear Regression Results

Table K1

Bivariate Linear Regression Coefficients for Predictors and Adoption

Independent		Unstandardized. Coefficients		Standard. Coefficients			95% CI for B	
Variable [a]	Const	*b*	Std. Error	Beta	*t*	Sig.	Lower	Upper
Rel. advantage	3.81	.083	.125	.052	.664	.507	-1.64	.331
Compatibility	3.28	.176	.119	.116	1.48	.139	-.058	.441
Complexity	8.15	-1.068	.154	-.478	-6.94	.000	-1.37	-.764
Observability	2.51	.358	.139	.198	2.57	.011	.083	.633
Trialability	5.94	-.298	.119	-.192	-2.49	.014	-.534	-.062

Note. [a] Criterion variable: Adoption

Table K2

Bivariate Linear Regression Model Summary for Predictors and Adoption

Independent		R	Change statistics [b]		Durbin-
Variable [a]	R	Square	F Change	Sig. F chg.	Watson
Relative advantage	.052	.003	.441	.507	1.974
Compatibility	.116	.013	2.205	.139	1.962
Complexity	.478	.228	48.219	.000	2.011
Observability	.198	.039	6.616	.011	1.963
Trialability	.192	.037	6.220	.014	1.983

Note. [a] Criterion variable: Adoption. B df1 = 1; df2 = 163