2017

Proactive IT Incident Prevention: Using Data Analytics to Reduce Service Interruptions

Mark G. Malley
Walden University

Walden University

College of Management and Technology

This is to certify that the doctoral study by

Mark Malley

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. Kelly Chermack, Committee Chairperson, Doctor of Business Administration Faculty

Dr. Natalie Casale, Committee Member, Doctor of Business Administration Faculty

Dr. Gail Ferreira, University Reviewer, Doctor of Business Administration Faculty

Chief Academic Officer
Eric Riedel, Ph.D.

Walden University
2017

Abstract

Proactive IT Incident Prevention: Using Data Analytics to Reduce Service Interruptions

by

Mark Gregory Malley

MS, Drexel University, 2011

BS, Drexel University, 2003

Doctoral Study Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Business Administration

Walden University

February 2017

Abstract

The cost of resolving user requests for IT assistance rises annually. Researchers have demonstrated that data warehouse analytic techniques can improve service, but they have not established the benefit of using global organizational data to reduce reported IT incidents. The purpose of this quantitative, quasi-experimental study was to examine the extent to which IT staff use of organizational knowledge generated from data warehouse analytical measures reduces the number of IT incidents over a 30-day period, as reported by global users of IT within an international pharmaceutical company headquartered in Germany. Organizational learning theory was used to approach the theorized relationship between organizational knowledge and user calls received. Archival data from an internal help desk ticketing system was the source of data, with access provided by the organization under study. The population for this study was all calls logged and linked to application systems registered in a configuration database, and the sample was the top 14 application systems with the highest call volume that were under the control of infrastructure management. Based on an analysis of the data using a split-plot ANOVA (SPANOVA) of Time 1, Time 2, treatment, and nontreatment data, there was a small reduction in calls in the number of reported IT incidents in the treatment group, though the reduction was not statistically significant. Implications for positive social change include reassigning employees to other tasks, rather than continuing efforts in this area, enabling employees to support alternative initiatives to drive the development of innovative therapies benefiting patients and improving employee satisfaction.

Proactive IT Incident Prevention: Using Data Analytics to Reduce Service Interruptions

by

Mark Gregory Malley

MS, Drexel University, 2011

BS, Drexel University, 2003

Doctoral Study Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Business Administration

Walden University

February 2017

Dedication

This study is dedicated to my family, friends, and coworkers who supported me along my educational journey by listening to my ideas and providing encouragement through my many years of study. Thank you!

Acknowledgments

While this doctoral study has a single author, its creation would not have been possible without the support of many individuals. I would first like to thank leadership within my organization for their support, especially Heinz Eppelmann for his input on researching the possibility of detecting incidents before they occur.

I am extremely thankful to have had such a talented chair, Dr. Kelly Chermack. Her knowledge, assistance, and feedback made completing this study as stress-free as possible. I also thank my second committee chair, Dr. Natalie Casale, for her time spent providing suggestions for strengthening my research, and Dr. Gail Ferreira, my university research reviewer, for ensuring my study met the standards of Walden University.

I would also like to thank the numerous faculty members at Walden University who prepared me for completing the research required to complete this capstone. Special thanks to Dr. Reggie Taylor and Dr. Freda Turner for their assistance in building the foundation of my study and supporting my committee.

I would also like to acknowledge the support of my doctoral cohort, who provided suggestions, feedback, and support during this journey.

Table of Contents

List of Tables

List of Figures

Section 1: Foundation of the Study

The use of information technology (IT) is a source of competitive advantage for organizations, though the complexities of IT often result in users requiring support to answer questions and resolve issues (Strohmeier, 2013). User confusion, frustration, and resistance can occur when organizations deploy new technologies, retire legacy systems, change or implement new infrastructure and processes, or offer alternative working styles (Rivard & Lapointe, 2012). Organizational leaders can provide internal help desks or purchase support offerings from third-party vendors to handle user requests for assistance, reducing the negative impact of technology use in the workplace (Ehrhart, Witt, Schneider, & Perry, 2011). Alternatively, organizational leaders can implement processes to reduce proactively the number of issues experienced by users, thereby reducing the number of support calls (Baskerville, Spagnoletti, & Kim, 2014).

Background of the Problem

The cost of resolving user requests for IT assistance rises annually, with some employers spending between $25-30 per incident to resolve employee IT support calls (Siti-Nabiha, Thum, & Sardana, 2012). The number of support calls per week is also rising, with large organizations receiving an average of 14,000 calls per week from users requiring assistance (Siti-Nabiha et al., 2012). Despite the increase in support requests, many IT leaders spend money to reduce the amount of time help desk staff devotes to resolving issues, instead of using resources to eliminate the cause of support calls (Bober, 2014). Employees working on an effective service desk typically resolve incidents quickly enough to meet a service level agreement (SLA), and do not spend time reducing the number of incidents created (Richardson & Mahfouz, 2014).

According to Moustaghfir (2012), in many organizations, the knowledge to prevent support calls already exist internally, but IT staff does not act on existing knowledge to proactively prevent incidents. The possibility exists that using less knowledgeable support staff to respond to users' requests for assistance is more cost effective than using highly paid and highly knowledgeable staff to prevent problems proactively (Boronico, Zirkler, & Siegel, 2011; Richardson & Mahfouz, 2014). However, IT leaders may not be aware of whether their organizations could cost effectively use existing organizational knowledge generated from data warehouse analytical measures to reduce the number of IT incidents (Provost & Fawcett, 2013).

Problem Statement

Large organizations can spend $350,000 to $420,000 per week resolving IT support calls, with some large organizations handling an average of 14,000 calls per week (Siti-Nabiha, Thum, & Sardana, 2012). Yet in many organizations, up to 90% of all IT support calls are preventable through employee use of existing organizational knowledge (Moustaghfir, 2012). The general business problem is that some employees within IT organizations create processes and implement knowledge management systems, but often may not analyze internal data to prevent future incidents. The specific business problem is that managers of an international pharmaceutical company, with headquarters in Germany, do not know whether using organizational knowledge generated from data warehouse analytical measures globally reduces the number of reported IT incidents.

Purpose Statement

The purpose of this quantitative, quasi-experimental study was to examine whether information technology (IT) staff use of organizational knowledge generated from data warehouse analytical measures reduced the number of IT incidents reported by global users of IT within an international pharmaceutical company headquartered in Germany. There were three independent variables in this study. The first independent variable was *application*, which represented the application system name. I analyzed 14 application systems in this study: seven in a control group, and seven in a treatment group. The second independent variable was *Time 1*, which consisted of the number of support calls received during an initial 30-day period before applying data warehouse analytical measures to the treatment. The third independent variable was *treatment*, which consisted of the application of data warehouse analytical measures. The dependent variable was *Time 2*, consisting of the number of support calls received for all applications during a 30-day period when the treatment group had data analytical measures applied. This study's implications for positive social change included potential contributions of new knowledge and insights which may lead to financial savings for leaders within the pharmaceutical industry, and the ability of employees to spend more time developing innovative therapies benefiting patients.

Nature of the Study

I used a quantitative methodology for this study, though Goertz and Mahoney (2012) stated one of three research methods (qualitative, quantitative, or mixed methods) are appropriate for researchers to use to when designing and conducting studies.

Researchers choose quantitative designs to examine whether a significant relationship exists between two or more variables (Applebaum, 2012; Goertz & Mahoney, 2012). For instance, Boronico et al. (2011) and Järveläinen (2013) used quantitative designs to examine the relationships between incident resolution, costs, and incident prevention. Given that I sought to examine similar relationships, this method was appropriate for this study. A qualitative method was not appropriate for this study as I was interested in understanding the relationship between variables, and not the lived experiences, actions, or motivations of users' in the study (see Oun & Bach, 2014). A mixed-methods study is a combination of qualitative and quantitative methods, and was not appropriate for this study due to time constraints in gathering data and challenges obtaining objective data from users (Fassinger & Morrow, 2013). In addition, there was not a need for a qualitative, employee perspective analysis to supplement the quantitative data of this study. My goal was to examine the effect of a treatment using data warehouse analytic measures.

The design of this study was quasi-experimental, a design used by researchers performing mixed, between-within subjects ANOVA, also known as a split-plot ANOVA or SPANOVA (Edmonds & Kennedy, 2012; Pallant, 2013; Tabachnick & Fidell, 2013). This design was appropriate because of the availability of archival data from a help desk ticketing system. Sheikh, Nurmatov, Cresswell, and Bates (2013) proposed using a quasi-experimental design to understand how best to deploy limited resources while examining the relationship between cost-effectiveness and healthcare information technology. Alternative designs include descriptive research, correlational research, and experimental research (Gravetter & Forzano, 2015). Edmonds and Kennedy (2012) indicated that descriptive research is most suitable for researchers seeking to describe the current status of variables, which was not applicable to the research question of this study. According to Breen, Holm, and Karlson (2014), researchers choose correlational designs to understand the scope of variable relationships when they are not manipulating an environment. A correlational design was not suitable for this study because organizational leaders desired to learn the impact of implementing data warehouse analytic techniques, a form of manipulation (see Breen, Holm, & Karlson, 2014). Researchers employ true experimental designs when establishing the cause and effect relationship among variables (Cooper & Schindler, 2013). The possibility did not exist for me to randomly assign users to groups to conduct a true experimental design for this study.

Research Question

My overarching research question was: Does using organizational knowledge generated from data warehouse analytical measures reduce the number of reported IT incidents at an international pharmaceutical company headquartered in Germany? My specific research question was: Does the use of data warehouse analytic measures (treatment) in a treatment group (Time 2) create a statistically significant reduction in the number of reported IT incidents among applications (application) over a 30-day period, when compared to a group of seven additional applications used as a control group (Time 1)?

Hypotheses

To answer the research question, I generated hypotheses to examine the relationship between the research variables of (a) application (application system name), (b) Time 1 (support calls during Time 1), (c) Time 2 (support calls during Time 2), and (d) treatment (application of the treatment). The hypotheses were:

$H1_0$: There is no significant difference in the number of incidents reported during Time 1 between treatment and nontreatment applications.

$H1_a$: There is a significant difference in the number of incidents reported during Time 1 between treatment and nontreatment applications.

$H2_0$: There is no significant difference in the number of incidents reported during Time 2 between treatment and nontreatment applications.

$H2_a$: There is a significant difference in the number of incidents reported during Time 2 between treatment and nontreatment applications.

$H3_0$: There is no significant difference in the number of incidents reported in the treatment application group during Time 1 and Time 2.

$H3_a$: There is a significant difference in the number of incidents reported in the treatment application group during Time 1 and Time 2.

$H4_0$: There is no significant difference in the number of incidents reported in the nontreatment application group during Time 1 and Time 2.

$H4_a$: There is a significant difference in the number of incidents reported in the nontreatment application group during Time 1 and Time 2.

Theoretical Framework

I used organizational learning theory in this study. Argyris and Schön (1978) developed this theory as a way to describe how members of organizations learn from previous mistakes to arrive at a different outcome. Dodgson (1993) further developed organizational learning theory as an explanation for how members of organizations learn, first by conceiving an idea, then executing on the idea, and finally reflecting on the actual results. Argyris and Schön (1978) described three key constructs for organizational learning: single-loop learning, double-loop learning, and deuteron-learning.

Single-loop learning involves making small changes or corrections to resolve a problem or improve operations (Argyris and Schön, 1978; Baskerville et al., 2014). Baskerville et al. (2014) stated that the single-loop form of learning occurs when people analyze specific actions instead of the greater context of measures taken, or the broader governing principles. Double-loop learning occurs when members of an organization go beyond the error-detection-correction cycle and prevent future errors by changing core organizational processes, assumptions, and beliefs (Argyris & Schön, 1978; Dodgson 1993).

Deutero-learning, or triple-loop learning, is a type of transformational learning that occurs when organizational members reflect on how they learn (Argyris, 1994). According to Nevis, DiBella, and Gould (1995), when employees realize they lack knowledge and implement policies and procedures to obtain missing knowledge, they are executing deutero-learning. Organizational learning theory was applicable to this study because organizational leaders wanted to learn how to prevent future issues from occurring by leveraging existing organizational data.

Definition of Terms

Many terms used throughout this study may not be familiar to readers, while numerous names exist for similar concepts. The following are operational definitions of the terminology I used in this study, provided to resolve any ambiguity.

Analytical measures: Machine learning and statistical techniques applied to the data so organizational members can gain insights to improve business processes (Boire, 2015).

Change management: A method of documenting modifications to an environment along with processes for review and approval by stakeholders. This method is also referred to as change control (Babalâc & Uda, 2014; Sykes, 2015).

Data warehouse: A collection of structured and unstructured data gathered from business processes stored in a database management system (Oktavia, 2014)

Service desk: Also referred to as a support desk or help desk, this is a team responsible for gathering requests and notifications (Jaiswal & Levina, 2012; Lund, 2012). This team resolves requests when possible, and routes the requests to others as necessary (Jaiswal & Levina, 2012; Lund, 2012).

Support call: An official record of an issue reported to a central service desk by a user or machine, which is stored in a database for informational and actionable purposes, also referred to as a service call or ticket (Datta & Lade, 2015; Son, Hazlewood, & Peterson, 2014).

Assumptions, Limitations, and Delimitations

Assumptions

Assumptions in a research study are pieces of information assumed to be true in order to test a theory (Foss & Hallberg, 2014; Simon & Goes, 2013). My first assumption was that the sampled systems were representative of an entire environment. Because the organization uses hundreds of application system environments, analyzing each system that was part of this study was not possible. To obtain adequate data, I analyzed the 14 application systems under the control of the IT infrastructure organization with the most support calls.

My next assumption was that each 30-day time period would include the typical number of support calls. Thirty days covers over 4 weeks, or 720 hours. A possibility exists that some applications in the environment were only used at certain times throughout the year, such as during the end of a quarter or at the end of the year. Applications not used regularly could have resulted in support calls logged during one 30-day time period but not another, distorting the effectiveness of any intervention with the results reflecting only a snapshot in time (Salthouse, 2011). To mitigate the risk of capturing data from sporadically used applications, I reached out to the lead for each application selected for the study and determined when the employees typically use the application. I also randomly selected seven applications to comprise a control group, which did not receive treatment.

My third assumption was that users would report incidents they experience with applications to the global service desk. When users experienced an issue, they may have searched the Internet for a solution, asked a colleague for help, or directly contacted the person responsible for the application. If users bypassed the service desk, service desk employees would not have capture the incident information for further analysis. To mitigate the risk of missing information about issues, regular reminders were sent to users asking them to use official communication channels for reporting issues. Additionally, according to an internal process document, application system leads received instructions to open up an official call when contacted directly by users.

My final assumption was that the data warehouse analytical measures would be robust enough to generate useful data that IT professionals can use to prevent future incidents. Employees analyzing call data might have gained no useful insights when analyzing past calls or log file data, which would have prevented the application of any treatment to the application groups. To mitigate this risk I used industry standard analytic tools and followed vendor supplied best practices.

Limitations

Limitations result from design or methodology influences outside of researcher control that have the potential to impact the research findings (Ekekwe, 2013; Simon & Goes, 2013). The following limitations existed because of the nature of the study, the sample population, limited resources, and time constraints. To determine if the application of data warehouse analytical measures reduced the number of IT incidents, I needed to determine the number of tickets opened before the application of the data warehouse analytical measures. IT staff could only analyze incidents reported to the global service desk and appropriately documented through the service desk ticketing system. Incidents not captured by the service desk were not a part of the analysis, which might have caused erroneous rejection of a hypothesis.

When IT staff members received notification of an end user issue, other staff members accessing the help desk ticketing system would only learn of the outcome of the end user issue if the staff member resolving the issue properly updated the ticket in the ticketing system. The assignment of a valid resolution code was the only way to capture and report on reported incidents. Common resolution codes included: (a) issue resolved, (b) known issue, (c) the user is unreachable, (d) workaround provided, and (e) no resolution possible. The stated goal of IT managers is for IT employees to resolve tickets as quickly as possible and with a status of resolved. If the ticket resolution information was not accurate or complete, employees could face difficulty when attempting to apply data warehouse analytical measures to the data, leading to incorrect conclusions.

The data in this study only applied to one international pharmaceutical company's users and applications. Because thousands of application systems are used within the company, it was not possible for me to examine all of the application systems used by employees, nor was it possible to test out every tool capable of performing data analytics. With over 45,000 users of IT, I could not follow-up on every support ticket opened by users within the organization.

Delimitations

Delimitations are design or methodology decisions within the control of researchers, set to maintain the manageability of the study (Simon & Goes, 2013). The first delimitation I established was that the data for this study would come from IT service calls made during two 30-day periods by internal and external employees in an international pharmaceutical company with over 45,000 employees. A longer duration would have caused delays in completing the study within a reasonable timeframe. My use of archival data from a database of service calls instead of direct contact with employees may have limited my ability to create a complete picture of IT issues.

Another delimitation was that the data I would analyze would come from calls related to the top 14 applications under the control of the infrastructure tower within IT in use at the company. The chosen applications were those most commonly used by end users, along with those most critical to the functioning of the business. Analyzing the call data from additional applications would have delayed the completion of the study without a guarantee of providing additional insights. I selected the company because of the ease of access to the data, and the support of the organizational leaders to implement a data warehouse analytic technique treatment.

Significance of the Study

Employees within organizations in multiple industries use data analytic tools including: (a) banking, (b) social media, (c) college recruitment and retention, (d) crime prevention, (e) insurance modeling, (f) warranty claims processing, (g) talent acquisition, (h) and e-commerce (Burdon & Harpur, 2014; Greco & Aiss, 2015; Jinka et al., 2012; Moses & Chan, 2014). The use cases for predictive analytical tools include: (a) risk management, (b) enhanced capital allocation, (c) root cause identification, (d) improved end-user experiences, and (e) quicker decision making (Blossom, 2014; Burdon & Harpur, 2014; Halladay, 2013). According to Sykes (2015), organizations with widespread data analytic tool usage have large and mature IT departments to support the tools. Despite the role IT staff plays in implementing data analytic tools, information does not exist regarding the benefit of using predictive analytic tools within IT departments as a way to reduce service interruptions.

There are many advantages to using predictive analytics, but there are also risks including the release of confidential information and reduced employee productivity (Alabi, Beckman, Dark, & Springer, 2015; Domingos, 2012). The significance of this study is in determining if a statistically significant reduction of IT incidents follows the implementation of data warehouse analytical measures in a global pharmaceutical company headquartered in Germany. Only IT managers can determine if any reduction in reported incidents justifies the costs of implementing and supporting the data warehouse analytic measures.

Contribution to Business Practice

IT is a critical tool business leaders rely on to support business functions, and service disruptions from faulty IT systems often impact profitability (González, Giachetti, & Ramirez, 2005). Because of limited resources, IT workers have to balance the time and money spent resolving issues against time spent proactively preventing issues (Boronico, Zirkler, & Siegel, 2011). This study fills a gap in the scholarly and professional understanding of the better use of IT resources within organizations, specifically in the area of incident resolution versus incident prevention through the use of data warehouse analytical measures.

Organizational leaders have access to vast amounts of end-user support call information but do not know if it is possible to improve service offerings by analyzing this data (Bivall & Mäkitalo, 2013). Organizational leaders also are unaware of whether applying organizational knowledge generated from data warehouse analytical measures prevents future IT systems interruptions for employees (González et al., 2005). The findings of this study are of value to business leaders wishing to know if investing in data warehouse analytical measures will reduce the number of IT incidents. Depending on the findings, business leaders may alter the way they structure technology teams and invest in IT systems.

Implications for Social Change

A program put in place to enhance the availability and integrity of IT systems containing data relevant to patient safety ultimately benefits patients and increases profits (Armando et al., 2015). Employee job satisfaction could increase when end-users of technology do not face issues, and IT workers can feel more job satisfaction when time is spent developing technical solutions instead of resolving issues (Siti-Nabiha et al., 2012). End-users could be less resistant to adopting new technologies when the technology works as expected, allowing for the rapid and successful rollout of IT systems (Rivard & Lapointe, 2012).

Many systems within pharmaceutical companies are in place to ensure patient safety. These systems range from adverse reporting databases, systems housing formulary data, and personally identifiable healthcare information (Armando, Bezzi, Metoui, & Sabetta, 2015). The release of private data, the unavailability of databases, or the corruption of data within systems can result in significant fines from regulatory agencies and potential harm to patients (Armando et al., 2015). This study's implications for positive social change included potential contributions of new knowledge and insights which may lead to financial savings for leaders within the pharmaceutical industry, and the ability of employees to spend more time developing innovative therapies benefiting patients.

A Review of the Professional and Academic Literature

The purpose of this quantitative, quasi-experimental study was to examine whether IT staff use of organizational knowledge generated from data warehouse analytical measures reduced the number of IT incidents reported by global users of IT within an international pharmaceutical company headquartered in Germany. To gain insights on the topic, I performed a review of available peer-reviewed sources including: (a) conference proceedings, (b) doctoral dissertations, (c) management journals, (d) academic journals, and (e) periodicals. My thorough assessment of existing research related to the study topic served as the foundation of this study (see Onwuegbuzie, Leech, & Collins, 2012). I made limited use of sources that were not peer-reviewed, including them only where necessary to provide a complete review of the current state of the industry.

I began the literature review by generating search terms related to the topic *predictive analytics*. My strategy for the literature review was to identify and review scholarly sources relating to predictive analytics and incident prevention. The search terms included: (a) *predictive analytics*, (b) *data analytics*, (c) *big data*, (d) *end-user* *support*, (e) *data warehouse*, (f) *knowledge management*, (g) *predictive help desk*, (h) *preventing help desk calls*, (i) *incident prevention*, (j) *Hadoop*, (k) *business intelligence*, and (l) *data mining*. I gathered the literature used in this review from searches I conducted of multiple electronic databases such as Association for Computing Machinery (ACM), EBSCOhost, Google Scholar, ProQuest Central, and ProQuest Dissertations and Thesis.

Searches of the ACM database resulted in the most hits: 849 for *predictive analytics*, 1,074 for *help desk*, 35 for *help desk analytics*, 93 for *help desk knowledge management*, 155 for *data warehouse analytic measures*, 370 for *incident prevention*, 34 for *incident prevention predictive*, and 196 for *Hadoop predictive*. Many of the results were conference proceedings lacking peer-review, while other results did not relate to the research topic, leading to additional searches on EBSCOhost, Google Scholar, and ProQuest Central. The search terms *knowledge management, data analytics*, and *Hadoop help desk* were too broad and thus not useful.

An EBSCOhost search of peer-reviewed sources published from 2012 to 2015 produced in 44 results for the term *predictive analytics* and 12 for the term *help desk*. There were nine sources related to the term *service desk*, and an additional nine sources related to *end user support*, though none were relevant to the topic. A ProQuest search of peer-reviewed sources from 2012 to 2015 produced 54 results for the term *predictive analytics*, 46 for the term *help desk*, and 59 for the term *support desk.* Searches for the terms *big data, data analytics desk, data analysis support, data mining, end user support,* *data warehouse, incident prevention, knowledge management, preventing help desk calls* and *service desk* resulted in duplicate or irrelevant topics.

The study contains 147 sources of which 127 (86%) are dissertations, peer-reviewed journals articles, government sources, or seminal sources, and 138 (94%) of which have been published within 5 years of my anticipated date of graduation. In this literature review, I reference 109 dissertations, peer-reviewed journals, government sources, or seminal sources, 105 of which have were published since 2011 (96%). Listed in Table 1 are the types and publication dates of the resources I used in the literature review.

Table 1

Literature Review

Resources	Recent references (within 5 years of anticipated graduation)	References older than 5 years	Total
Books	3	1	4
Dissertations	1	0	1
Peer-reviewed articles	97	2	99
Other resources	4	1	5
Column total	105	4	109
Percentage of total sources used in this lit review	96%	4%	100%

This literature review includes a description of the general research in the area of data analytics and reducing help desk support calls, both qualitative and quantitative, as well as a discussion of and justification for my research approach and design. The 10 sections of the literature review include reviews of the literature on: (a) research design and method (b) the organizational learning theoretical framework, (c) the background of corporate IT, (d) organizational knowledge and knowledge management systems, (e) help desks and IT service desk processes, (f) methods for reducing calls to help desks, (g) current uses of predictive analytic measures, (h) predictive analytics tools, and (i) the research variables I used in the study.

Organizational Learning Theoretical Framework

I used organizational learning theory as the theoretical framework for this study. Organizational learning theory, as defined by Nevis et al. (1995), is a framework that organizational members can use to improve performance based on prior experience. The theory was first developed by Argyris and Schön (1978) as a way to describe how members comprising organizations learn from previous mistakes to subsequently achieve different results. According to Dodgson (1993), Argyris and Schön offered the theory as an explanation of how organizational members learn first by conceiving an idea, then executing on the idea, and finally reflecting on the actual results. Argyris and Schön described three key constructs for organizational learning: single-loop learning, double-loop learning, and deutero-learning.

Single-loop learning involves making small changes or corrections to resolve a problem or improve operations (Baskerville et al., 2014). According to Baskerville et al. (2014), an example of single-loop learning is when employees take corrective action to resolve an already occurring issue, but do not learn from the experience and will take the same corrective action if the issue reoccurs. An example of single-loop learning in an IT context is staff reviewing the user calls from the previous week to determine if a problem exists with specific systems, and then implementing a fix to eliminate the issue and restore functionality to reduce support calls from users.

Double-loop learning occurs when members of an organization go beyond the error-detection-correction cycle and prevent future errors by changing core organizational processes, assumptions, and beliefs (Argyris & Schön, 1978; Dodgson 1993). Baskerville et al. (2014) stated that employees exhibiting double-loop learning would gain insights into why a fix led to resolving an issue, potentially eliminating future occurrences of the issue. Many service desks operate on the principles of single- and double-loop learning (Moe, Šmite, Hanssen, & Barney, 2014). When a user experiences an issue, such as a forgotten password, the service desk representative follows established procedures to resolve the issue, such as resetting the password. When the number of calls exceeds an established benchmark, changes to the underlying systems can be made, such as implementing a forgotten password feature in the program, to permit the user to remember the password and continue working (Siti-Nabiha et al., 2012). The owner of the application could also implement alternative means for users to authenticate themselves, such as fingerprints that users cannot forget (Kumari, Gupta, Khan, & Li, 2014).

Deutero-learning, or triple-loop learning, is a type of transformational learning which occurs when organizational members reflect on how they learn (Argyris, 1994). According to Nevis et al. (1995), when employees realize they lack knowledge and implement policies and procedures to obtain knowledge, they are executing deutero-learning. Organizational learning theory, and specifically deutero-learning, is pertinent for this study because when organizational managers realize they lack specific knowledge on how to prevent future issues from occurring, they want to ascertain if they can reduce the number of reported IT incidents by leveraging existing organizational data. Organizational leaders desire to learn how to proactively prevent incidents by fundamentally changing the assumptions surrounding the reporting and resolution of IT issues by users of IT application systems. Firms with strong organizational learning can have higher incidents of innovative behavior by employees, and employees in companies with strong organizational learning practices adapt quicker to external changes than employees within companies with little or no organizational learning capabilities (Wu, 2016).

Many researchers have used the organizational learning framework to gain insights on improving IT service offerings (García-Morales, Jiménez-Barrionuevo, & Gutiérrez-Gutiérrez, 2012; Moe et al., 2014; Noruzy, Dalfard, Azhdari, Nazari-Shirkouhi, & Rezazadeh, 2013). Moe et al. (2014) used organizational learning theory to explore why offshore IT outsourcing fails at small- and medium-sized companies, while García-Morales et al. (2012) and Noruzy et al. (2013) demonstrated how transformational leadership and organizational innovation impact organizational learning and knowledge management. Organizational learning theory is only one explanation of how IT investments create and sustain competitive advantages within companies (Mithas, Tafti, Bardhan, & Goh, 2012).

Background on Corporate IT

Users required support for data processing before the advent of end-user computer approximately 30 years ago, but advances in technology that enabled cost-effective and easy-to-use software spurred the creation of the large IT departments within organizations (Govindarajulu, 2014). Beginning in the 1960s, it was possible to use machines to store, retrieve, transmit, and manipulate massive amounts data to gain a competitive advantage and improve service (Frizzo-Barker & Chow-White, 2014). The people and systems that support data processing comprise organizational IT (Schäfferling & Wagner, 2013).

Modern IT departments provide support for commonly used tools including databases, web services, storage, applications, end-user devices, networks, telephones, instant messaging, email, and servers (Bartolini et al., 2011). The size of the IT department is usually dependent on the size of the organization (Datta & Lade, 2013). In a small business, the IT department can consist of a single technologically savvy employee (Datta & Lade, 2013). In large organizations, an IT department consists of many smaller and highly specialized teams responsible for delivering services (Bartolini et al., 2011). Small specialized IT teams work together transparently to provide technology services expected by end users (Li, Wu, Yen, & Lee, 2011).

Due to the demand from business users for feature-rich application systems, IT staff members are under intense pressure to deliver productive systems that remain available to users (Allman, 2012; Li et al., 2011). Often users experience delays in gaining access to new technologies, and when users do obtain access, they may experience performance issues, bugs, missing features, and cryptic error messages (Allman, 2012). Because most IT departments do not directly generate profit for organizations, IT managers, including chief information officers, often face resistance from organizational leaders to hiring enough staff to break the cycle of supporting poorly executed IT systems. Help desk staff often bear the brunt of the adverse effects (Allman, 2012; Armstrong, Simer, & Spaniol, 2011; Brown & Brandt, 2014).

Many organizational leaders consider IT an *enabling function*, a department that discovers new business opportunities to drive growth, similar to a human resources department, where employees seek out innovative ways to identify new talent to increase profits (Brown & Brandt, 2014; Kornberger, 2016). Despite delivering innovative services, IT as an enabling function is not a significant generator of revenue for organizations, and therefore members of IT are under pressure to deliver reliable services as inexpensively as possible (Brown & Brandt, 2014). Depending on the views of senior management, IT can be supplemented by or sourced entirely from an outside service provider in an attempt to reduce costs, enhance and increase service offerings, or improve execution times (Armstrong et al., 2011).

Organizational Knowledge and Knowledge Management Systems

Institutional knowledge is the combined knowledge of individual contributors within organizations, and is used collectively by employees to gain a competitive advantage over rival firms (von Krogh, Nonaka, & Rechsteiner, 2012). Also called, organizational knowledge, institutional knowledge includes any knowledge employees can leverage to gain efficiencies over competitors, not only knowledge related to the core business (von Krogh et al., 2012). Within the pharmaceutical industry, organizational knowledge generated within IT departments can provide firms with significant advantages by reducing drug discovery times, reducing fines by maintaining compliance with regulations, and enable knowledge workers to focus on their core jobs without distractions from underlying technologies (von Krogh et al., 2012).

Sharing of organizational knowledge between groups benefits scientific research, though effective tools are necessary to support productive sharing (Chong, Skalka, & Vaughan, 2015). Many researchers refer to knowledge management as knowledge sharing, as the ultimate goal behind knowledge management system implementations is to facilitate the sharing of knowledge throughout the organization (Krogh et al., 2012; Mithas et al., 2012; Sousa, Costa, & Aparicio, 2013). According to Sousa et al. (2013), employees need to understand that organizational knowledge is not always accurate or useful knowledge. Nazari and Emami (2012) noted approximately 70% of organizational knowledge management system implementations fail, often because users do not see any benefit in contributing knowledge to or retrieving knowledge from the system, or the systems are too complicated to use.

One challenge with sharing data throughout organizations is the sheer volume of information generated, and the lack of tools and employee expertise to analyze the data and extract actionable information (Borkar, Carey, & Li 2012; Krogh et al., 2012; Sousa et al., 2013). The design of early knowledge management systems had a file system with a structure similar to physical filing cabinets, but as the volume of data increased, organizations implemented enterprise databases capable of supporting large relational queries (Borkar et al., 2012).

Due to the relatively low cost of storage, recording vast amounts of data generated by organizations became possible and is a routine practice within many organizations, but information overload can lead to frustration, anxiety, and poor decision making (Green 2016). Organizations often err on the side of caution and record more data than is necessary for any business purpose since employees find capturing and storing data an easy task to perform (Blossom, 2014). The challenge members of organizations face include deriving value from the information organizations possess, and transforming the data into knowledge (Sousa et al., 2013). The required processing power of computers to analyze volumes of data increases exponentially as database sizes increase, further hindering the ability of organizational leaders to implement effective knowledge management systems (Borkar et al., 2012).

To address the demand for better tools to process data, software developers at large web firms, including as Google, Facebook, and Yahoo!, created proprietary file systems spanning multiple machines to process information (Borkar et al., 2012). Alternatives soon appeared on the market, including: (a) Hadoop from Apache, (b) Dryad from Microsoft, and (c) Jaql from IBM (Borkar et al., 2012). Apache's Hadoop is a widely used open-source tool, used to performed long analysis and website indexing (Borkar et al., 2012; Ordonez, 2013).

Ordonez (2013) argued that while Hadoop is the leading technology for analyzing big data, traditional database management systems (DBMS) remain effective tools for processing large data sets. The choice of tool depends on the complexity and size of the data set, available IT budget, and the skills and familiarity IT professionals have with the products on the market (Ordonez, 2013). Organizations with complex data processing needs could use a variety of tools to analyze big data, and the choice of tool often depends on the department using the tool (Ordonez, 2013).

Help desk software is a type of knowledge management system, a software product help desk members use to decrease the time to resolve user issues (Sousa et al., 2013). When a user calls, information is available to the agent from the help desk software interface, including: (a) user location, (b) prior call history of the user, (c) prior incidents reported regarding the affected system, and (d) next steps for the agent to take to begin resolving the user's call (Edwards, 2015). Organizations primarily implement help desk knowledge management systems to decrease the time to resolve issues, though help desk software can perform additional functions (Katzan, 2011). By resolving user issues quickly so employees can focus on their core jobs, help desk workers can provide organizations a competitive advantage (Katzan, 2011).

Despite the competitive advantage provided by employees sharing knowledge, in many organizations, managers can face difficulty motivating employees to share information in traditional knowledge management systems (Wang, Noe, & Wang, 2014). Ackerman, Dachtera, Pipek, and Wulf (2013) noted that information-sharing problems rarely exist with support systems, as users are willing to provide as much information as necessary to solve the problem, often without privacy concerns. Employees can be too willing to share information to resolve problems, not understanding the consequences of sharing personal information, and violating company policies and security best practices by sharing passwords and allowing others access to confidential information (Barlow, Warkentin, Ormond, & Dennis, 2013).

Despite employee willingness to share information informally with colleagues, organizational leaders often encounter employee resistance when trying to implement knowledge management systems (Wang et al., 2014). Challenges with creating effective knowledge management systems include: (a) employee willingness to contribute data, (b) determining the data to capture within the system, (c) maintaining data within the system, (d) training employees to use the system, (e) cost of implementing the system, and (f) data accuracy (Ackerman et al., 2013; Nazari & Emami 2012). IT management supports the creation of knowledge management systems as a way to cost-effectively fulfill end-user requirements (Wang et al., 2014).

The deployment of help desks is the most common response to fulfilling end-user support requirements, though the possibility exists for end-users to receive faster resolutions with higher satisfaction by using wikis, forums, and user training (Leung & Lau, 2015). Leung and Lau (2015) stated that while implementing help desks is not the quickest way to resolve IT issues experienced by users, the use of help desks is often more cost effective than using highly paid peers and IT specialists to resolve common IT issues. Help desk staff members themselves often use knowledge management systems to resolve user support calls (Leung & Lau, 2015).

Help Desks and IT Service Desk Processes

The roles of help desks within organizations are to serve as a central resource for users to contact when they require assistance (Siti-Nabiha et al., 2012). Help desk employees can provide guidance on a broad range of topics, including technical issues with computer applications and technology systems, misplaced identification card, facility issues, and problems with pay or benefits (Herrick et al., 2012). As the role of help desk employees expanded, help desks also became known as *service desks*, reflecting the role of help desk employees of not only providing assistance for problems but routing and fulfilling requests (Siti-Nabiha et al., 2012). Users can typically access an organization's help desk by telephone, email, instant message, text message, web forms, or a walk-up counter (Akinnuwesi et al., 2014).

The internal structure of IT departments and help desks varies based on organizational needs and cultural preferences (Kim, Lee, & Joshi 2013). The workers on a help desk may be internal to the organization, external to the organization, or a combination of both (Jaiswal & Levina, 2012; Robbins & Harrison, 2011). Help desk workers can support internal employees and external parties such as customers, suppliers, and government regulators (Lund, 2012).

Different tiers can exist within a help desk, with higher tier employees having more experience, knowledge, and responsibility (Datta & Lade, 2015). Higher-tier employees typically receive higher compensation; therefore, organizations strive to solve as many user issues as possible with first-tier support (Flynn & Philbin, 2014). Some help desk employees may only record user issues and forward the ticket to others for resolution (Akinnuwesi et al., 2014). Other organizations can enable help desk employees to resolve common problems, such as forgotten passwords and locked out accounts (Datta & Lade, 2013)

Though many help desk models exist, a successful help desk is one that is adequately funded and staffed while providing more value than the cost to provide the service (Ashcraft, Cummings, Fogle, & Valdez, 2015). According to Ashcraft et al. (2015), help desk employees provide value by eliminating the productivity loss from users being unable to complete assigned tasks due to issues using technology. Help desk managers typically employ enough staff members to maintain a queue depth of a predetermined length to balance speed of call resolution with the cost of idle workers (Robbins & Harrison, 2011).

When a user contacts a help desk, the first action is for the help desk agent to log the call in a database via the knowledge management system (Ma, Kim, & Rothrock, 2011). Help desk software exists to support help desk workers in collecting, categorizing, routing, and resolving problems (Son, Hazelwood, & Peterson, 2014). Help desk software is a tool help desk personnel can use to retrieve information about the user and the problem the caller is reporting (Datta & Lade, 2015). The help desk worker can review prior calls made by the user, determine if other users are experiencing the same issue, and notify the appropriate group about a potential outage (Herrick et al., 2012). Help desk workers function in a critical role for organizational leaders by serving as the first point of contact for users, and therefore being the first to discover potential outages (Siti-Nabiha et al., 2012). Help desks can be the sole point of contact between customers and a company; therefore, help desk employees influence a company's reputation (Dzuba, 2015).

Bartolini et al. (2011) stated that typical help desk tools in use by IT departments limit agility, as programmers designed the tools with a focus to capture data and not optimize IT resource deployment. As ticket call volume increases, senior leaders will deploy additional resources to resolve user issues and restore normal operations (Leung & Lau, 2015). While additional resources can result in quicker service restoration, IT managers typically do not have the money available to hire the staff necessary to resolve all issues as users call to report concerns (Herrick et al., 2012; Sykes, 2015). Software, such as data analytic tools and decision support tools, exist on the market, and the use of these tools can reduce user help desk calls as well as prioritize the order in which issues are resolved (Bartolini et al., 2011).

One reason for the existence of help desks is to quickly resolve issues, as unplanned outages result in lost business opportunities, user frustration, and potential compliance issues (Flynn & Philbin, 2014). Help desks employees can serve as a central resource for users to contact when users do not know who else to contact (Govindarajulu, 2014). While help desk staff members receive technical training on resolving common issues, in some settings 80% of calls to help desk personnel required no specialized technical training to resolve (Leung & Lau, 2015; Obasa & Salim, 2014).

While the goal of help desk operators is to support the timely resolution of user issues, IT managers see value in reducing the total number of calls placed to help desks, even if they cannot lower the number of user issues (Flynn & Philbin, 2014). Data available from resolved help desk calls can be beneficial to IT management attempting to assign a dollar amount to the cost of supporting applications, as well as facilitating data-driven decision making (Armstrong et al., 2011). Despite the wealth of information available from help desk call logs, Flynn and Philbin (2014) and Ashcraft et al. (2015) provided many reasons for reducing the number of calls to help desks, including: (a) reducing staff burnout, (b) preventing illness from stress, (c) reducing employee turnover, (d) eliminating caller demoralization from long wait times, and (e) changing negative employee perceptions resulting from poor customer service. IT leadership can employ multiple strategies to help reduce the number of calls help desk operators receive from users (Songsangyos, Niyomkha, & Tumthong, 2012).

Methods of Reducing Calls to Help Desks

Organizational leaders have many reasons for attempting to reduce the number of calls help desk employees receive. Operating help desks requires many organizational resources, and reducing the size of help desks along with the operating hours of help desks can potentially reduce costs (Songsangyos et al., 2012). Though Songsangyos et al. (2012) noted expert systems could reduce the number of calls placed to help desks, organizational leaders should not equate a reduction in user calls with reduced user issues or improved employee satisfaction with technology services. Despite the satisfaction that comes from the ability for help desk workers to resolve user issues, many IT professionals do not enjoy working at a help desk, and organizational employees dislike reaching out to help desk employees when problems occur (Jaiswal & Levina, 2012). One way to reduce employee complaints regarding working with help desk employees to resolve issues is through the implementation of automated resolution systems (Datta & Lade, 2015).

Datta and Lade (2015) demonstrated how organizations employing automated techniques could reduce the number of calls reaching help desks, though an automating help desk ticket process and providing poor customer service does not reduce the number of users with issues, only the number of calls processed by help desk workers. Dzuba (2015) noted that scripted responses from call center employees annoyed callers and reduced perceived service quality ratings. Providing a poor experience to users is a way to reduce help desk call volume, though at the expense of user productivity and satisfaction (Clark, Murfett, Rogers, & Ang, 2012). One solution to improve employee perceptions and satisfaction with IT departments is to eliminate the need for employees to seek out expert assistance (Datta & Lade, 2015).

Help desk employees could deploy automation to detect and repair problems before outages impact users (Asthana & Okumoto, 2012). According to Kaur and Kinger (2013), monitoring services can detect problems and automatically alert the responsible team or take corrective action without human intervention. One challenge with deploying automated systems is the time and resources necessary to deploy and maintain the systems (Songsangyos, Niyomkha, & Tumthong, 2012). To resolve problems, employees sometimes need to understand the root cause of issues, though the cost to determine the root cause can exceed the cost to remedy the issue (Selvaraj, Jayabal, Srinivasan, & Balasubramanie, 2015). Another problem with automation, identified by Vinueza (2013), was poor system implementation and configuration, leading to user confusion and frustration, ultimately wasting time and creating more work for users and IT staff.

As a compromise between preventing all problems and preventing zero problems, IT service providers can use their limited resources to build reports of the most common issues, and attempt to understand the root causes of those specific issues (Herrick et al., 2012). According to Herrick et al. (2012), common help desk tools generate reports on call metrics including: (a) the system causing the issue, (b) the user reporting the issue, (c) the time to resolve the issue, (d) the impact of the issue, (e) the number of issues assigned to a technician, (f) how many tickets are outstanding, and (g) the number of tickets resolved within the service level agreement (SLA). IT management can use these reports to recognize patterns in reported system issues and determine the best use of limited IT resources (Akinnuwesi et al., 2014).

Nakamura and Kijima (2011) created a framework to assist IT staff in reducing IT incidents, classifying incidents into six spaces: (a) failures of rationality, (b) failures of technology, (c) failures of behavior, (d) failures of evolution, (e) failures of structure, and (f) failures of regulation. The possibility exists that users are responsible for creating the problems they report to help desk employees, and users could be taught to recognize corrective actions they can take to prevent the issue from occurring in the future (Talla & Valverde, 2013). If users can recognize actions that lead to IT incidents, organizational leaders could implement programs to train users to take corrective actions before reportable issues occur, and break the unwanted incident management loop of the user experiencing an incident, calling the help desk, and the help desk member resolving the issue (Talla & Valverde, 2013; Tehrani & Mohamed, 2011).

Another option for reducing help desk calls is the creation of a self-service portal for IT users (Katzan, 2011). Self-service portals include wikis with information on how to resolve common issues, dashboards with current system status where users can determine if a problem is already known to IT workers, sites with frequently asked questions with answers, and tools to handle common problems such as resetting a forgotten password (Obasa & Salim, 2014). Both end-users and IT staff members can create the content stored within self-service knowledge management portals (Somvanshi, Gogate, Sahane, & Kolhe, 2014).

According to Obasa and Salim, (2014), multiple issues exist with self-service portals, such as the static nature of typical corporate portals. Additional issues with self-service portals include: (a) users could not know the self-service portal exists, (b) users not understanding how to access or use the system, or (c) users bypassing the system entirely, wishing instead to talk to a human (Obasa and Salim, 2014). A benefit of self-service portals is they are available 24x7 without human intervention, though they depend on IT systems being available to users, while help desk availability varies by the organization (Mithas et al., 2012).

Bartolini et al. (2011) demonstrated that another solution to the problem of increasing user calls is to implement decision support tools capable of simulating system performance after IT employees implement changes to computer environments. Decision support systems are capable of performing what-if scenario analysis, enabling IT workers to see the potential consequences of their actions (Bartolini et al., 2011). Such tools, if properly configured, assist IT workers in visualizing the links between systems (Bartolini et al., 2011). A worker could use a decision support system to learn that rebooting a specific server will bring the entire email system offline, and avoid rebooting the server during working hours to reduce the number of help desk calls by users reporting that they are unable to access their company email (Bartolini et al., 2011).

Creating peer support systems and using crowdsourcing are additional methods of reducing help desk calls, where employees are enabled and encouraged to share information with colleagues (Solemon, Ariffin, Din, & Anwar, 2013; Sykes, 2015). According to Sykes (2015), organizations with peer support systems typically have employees reporting higher IT and job satisfaction rates, less stress, and higher job performance than similar organizations without strong social networks. The use of crowdsourced knowledge is an effective method for employees to discover new ways to resolve issues and improve service offerings (Solemon et al., 2013). Sousa et al. (2013) confirmed the importance of socialization with the creation of a knowledge management conceptual model integrating organizational communication.

The reason peer support systems rank highly is that workers utilizing peer support systems bypass help desks, and users typically prefer to use the path with least resistance when resolving issues (Sykes, 2015). According to Sykes (2015), employees who query a peer typically do not have to provide background or irrelevant information, such as name and contact information. Employees using peer support systems also do need to wait on hold as they would when calling an overwhelmed help desk, or wait as a help desk employee routes the support ticket to the appropriate application group for resolution (Sykes, 2015).

Predictive analytic tools comprise a class of software that IT managers can implement to reduce the number of outages experienced by users, thereby reducing the number of calls received by help desk staff (Aggarwal, Bhagwan, De Carli, Padmanabhan, & Puttaswamy, 2011). The possibility exists for IT workers to implement software that could analyze the actions of a user, and detect if the user cannot log into an application, prompting the user to change a password and eliminating the need for the user to contact the help desk to report an issue logging into the application (French, 2015). While Aggarwal et al. (2011) demonstrated how some types of predictive analytic techniques could be non-disruptive and non-intrusive to end users, IT staff would still need to take action to prevent an outage from occurring and impacting users.

Predictive Analytic Measures

Predictive analytic measures are machine learning and statistical techniques applied to the data stored in data warehouses, implemented as a tool for organizational members to use to gain insights to predict and improve business processes (Boire, 2015; Halladay, 2013; Moses & Chan, 2014; Qureshi, 2014). Implementing machine learning algorithms is a cost-effective alternative to hiring staff for manual programming and data mining activities (Domingos, 2012; Parikh, Kakad, & Bates, 2016). Improved business processes from the implementation of predictive analytical measures include: (a) management of risk, (b) enhanced capital allocation, (c) root cause identification, (d) improved end-user experiences, and (e) quicker decision making (Blossom, 2014; Burdon & Harpur, 2014; Halladay, 2013).

Users of data analytics can make decisions based on statistical probabilities instead of hunches and intuition (Moses & Chan, 2014). Hayashi (2014) stated the value from using predictive analytics to save money is often clear to senior management, but successfully leveraging predictive analytic tools is difficult without having a clear understanding of the problem to solve, and the data necessary to solve the issue. Parikh et al. (2016) noted large organizations including Amazon and American Airlines, implemented predictive analytic tools, but the implementation of predictive analytics is not the only option available for gaining insights into ways of optimizing organizational practices. Other process improvement methods, such as ITIL, Six Sigma, and total system intervention for system failure are available for improving and optimizing organizational practices (Li et al., 2011; Nakamura & Kijima, 2011; Talla & Valverde, 2013; Tehrani & Mohamed, 2011; Valverde, Saadé, & Talla, 2014)

The data modeled by predictive analytic tools rests in data warehouses (Oktavia, 2014). According to Oktavia (2014), data warehouses are collections of structured and unstructured data gathered from business processes stored in a database management system. The data inside an IT data warehouse can come from many sources, including: (a) system log files, (b) user calls, and (c) monitoring software (Nagaraj, Killian, & Neville, 2012). Zliobaite et al. (2012) stated the raw data used to perform predictive analytics is typically not in a usable format; therefore, necessary to pre-process the data for analysis.

The varying sources of information within data warehouses create a challenge for IT workers attempting to implement predictive analytics (Schoenherr & Speier-Pero, 2015). Organizational data may not: (a) be available to IT staff, (b) exist in machine-readable formats, or (c) follow standardized data classification schemes (Burdon & Harpur, 2014; Qureshi, 2014; Zliobaite et al., 2012). Some types of confidential data, including patient records and human resources data, could not be available to IT staff due to government regulations, non-disclosure agreements, and company policies (Budgaga et al., 2016). Organizational data stored on password protected network file shares and data management systems could also be unavailable to IT staff (Budgaga et al., 2016). The possibility also exists that IT staff are not aware of the storage locations for organizational data (Richey, Morgan, Lindsey-Hall, & Adams, 2016). Though predictive analytic tools can benefit IT staff, usage of predictive analytic tools extends beyond the prevention of IT incidents (Burdon & Harpur, 2014; Moses & Chan, 2014; Prasad et al., 2015).

Organizational leaders face a paradox when collecting large amounts of data (McNely, 2012). According to Schoenherr and Speier-Pero (2015), many organizational leaders do nothing with the data employees collect, failing to leverage data with considerable value. Storing data can cost a significant amount as data are typically replicated to protect from the failure of the storage medium, backed up to protect against disasters, and archived to meet regulatory requirements (Beath, Becerra-Fernandez, Ross, & Short, 2012). Conversely, when organizational leaders process data, they must implement safeguards against the proliferation of misleading assumptions, which waste limited resources (Schoenherr & Speier-Pero, 2015). Motulsky (2015) noted that scientists often fail to analyze data at face value, and instead continue manipulating data until seeing a desired pattern or result, or until running out of resources to continue data analysis. Management should determine if the money spent performing predictive analytics would provide a greater return if spent elsewhere in the organization (Richey et al., 2016).

Predictive analytic tools have a wide range of uses, including: (a) banking, (b) social media, (c) college recruitment and retention, (d) crime prevention, (e) insurance modeling, (f) warranty claims processing, (g) talent acquisition, (h) e-commerce, and (i) supply chain management (Burdon & Harpur, 2014; Greco & Aiss, 2015; Jinka et al., 2012; Moses & Chan, 2014; Schoenherr & Speier-Pero, 2015). According to Jinka et al., (2012), predictive analytics are tools employees use to make better-informed decisions than otherwise possible with available data quickly. Outside of IT, employees use data analytics to predict the likelihood of a parolee becoming a repeat offender, the probability of a bank customer repaying a loan, or the chance a lawsuit will succeed (Moses & Chan, 2014). Users employing predictive analytics can learn by discovering patterns in data otherwise hidden by the huge volume of information available by using powerful computers (Blossom, 2014; Zliobaite et al., 2012).

Despite the many advantages to using predictive analytics, many risks also exist in using predictive analytics on data sets (Burdon & Harpur, 2014; Domingos, 2012). Burdon and Harpur (2014) noted a possibility exists to embed unwanted practices into the information infrastructure, leading to unintended discrimination in hiring. A possibility exists that a predicted scenario will never occur, but IT staff divert time and resources to implement preventative measures, never knowing the event would not occur if the IT staff took no action (Domingos, 2012). IT employees could enter a loop of executing preventative measures, each time thinking the measures prevented an outage when the outage was incorrectly predicted by the predictive analytic tool and would never have occurred (Domingos, 2012). The possibility also exists that any corrective action taken by IT members causes other unpredictable issues (Edwards, 2015).

Predictive analytic tools provide multiple methods to present relevant findings to operators, including: (a) pattern recognition, (b) rule-based engines, (c) data mining, (d) neural networks, and (e) data linkages (Jinka et al., 2012). Discovering that every third Monday a user account is locked out is an example of pattern recognition (Vera-Baquero, Colomo-Palacios, & Molloy, 2013). Rule-based engines apply logic to data and provide information to the operator (Tehrani & Mohamed, 2011). A rule-based engine may not flag a marketing email account sending out thousands of email messages a minute as compromised, but would to an account belonging to a member of the legal department (Salah & Chaudary, 2015).

Neural networks are a form of artificial intelligence, comprised of complex algorithms applied to data to perform predictive tasks (Shazmeen, Baig, & Pawar, 2013). A neural network could detect a denial of service attack is in progress and route users over a secondary network, preventing a future call about slow access times (Stevanovic, Vlajic, & An, 2013). Data linkages are associations between two or more items, with analytic tools processing data in a way similar to rule-based engines (Frizzo-Barker &

Chow-White, 2014). Users of data analytic tools could use some or all of these methods when providing data to operators (Jinka et al., 2012).

IT staff can use multiple predictive analytic methods to analyze massive amounts of data to determine patterns and predict issues (Halladay, 2013; Schoenherr & Speier-Pero, 2015). Zliobaite et al. (2012) stated that organizational leaders would find adaptive learning algorithms necessary to implement to handle the rapid growth of digital information. Although the benefits organizations obtain from implementing predictive analytic tools, data analytic tools remain in the domain of large organizations with mature IT departments (Sykes, 2015).

The predictive capabilities of data analytics are the differentiator from business intelligence, which is the use of historical information to influence business decisions (Halladay, 2013). According to Halladay (2013), organizational leaders use business intelligence to understand why something occurs, and use predictive analytics to understand what will happen *if* something occurs. The representative workflow for predictive analytics is to pre-process and transfer data to a central database, perform statistical modeling on the data, develop machine learning models, and finally apply the new model on future data (Alabi et al., 2015).

The rollout of predictive analytic tools could allow organizational leaders to respond quickly to employee needs, increasing employee satisfaction and productivity (Blossom, 2014). Organizational leaders discover interesting patterns when leveraging predictive analytics, though no researchers to date studied the impact of applying predictive analytics to help desks (Hayashi, 2014). Wal-Mart employees discovered during the days before a hurricane, demand for Pop-Tarts and flashlights increases (Hayashi, 2014). According to Hayashi (2014), when those responsible for purchasing decisions possess this type of information, they can increase sales while meeting customer expectations. Management in an international pharmaceutical company headquartered in Germany do not know what insights they will gain from implementing a predictive analytic tool within IT.

Predictive Analytic Tools

Despite advances in application and computer system development, users often experience unwanted and unexpected behaviors when interacting with technology (Ibidunmoye, Hernández-Rodriguez, & Elmroth, 2015). Predictive analytic tools used by IT professionals are software programs capable of performing machine learning and statistical techniques on data stored in a data warehouse used to predict a likely scenario or outcome (Boire, 2015). Data analytic tools can create predictions in real-time, as information is made available or generates predictions based on historical data (Amarasingham, Patzer, Huesch, Nguyen, & Xie, 2014). Due to the volume of information typically available from application system log files, most data analytic tools process historical data to generate the likelihood of a future system failure (Kim et al., 2013). IT staff can use predictive analytic tools to efficiently identify potential issues in modern and complex systems in use within organizations before the issues negatively impact users (Nagaraj et al., 2012).

Many challenges exist in implementing usable and trusted predictive systems (Zliobaite et al., 2012). Implementers need to identify sources of usable data to feed into the analytic tool (Kim et al., 2013). Domingos (2012) stated users of analytic tools will realize greater benefits as the amount of data made available to the tool increases, though Blossom (2014) disagreed, indicating the importance of carefully considering the data inputted into predictive analytic tools to reduce the possibility of noise from irrelevant data hiding actionable signals. In addition to data from log files, installers also need to integrate the organizational knowledge and expert knowledge into the system (Zliobaite et al., 2012). The complexity and scale of data analytic programs present another challenge (Nagaraj et al. 2012). HBase, the Hadoop data store, had over one million revisions from 2010 through 2012, and demonstrated the rapid pace of change within the field of predictive analytics (Nagaraj et al. 2012). While the tools used for performing predictive analytic are relatively immature, advances in capabilities of software and improvements in the performance of the hardware occur quickly (Kornberger, 2016).

Researchers are creating advanced tools to facilitate data analytics by non-experts, moving from adaptive algorithms to adaptive tools (Zliobaite et al., 2012). Nagaraj et al. (2012) noted IT professionals could leverage predictive analytic tools to quickly determine the root causes of issues, eliminating hours of manual labor. Burdon and Harpur (2014) cautioned that users of predictive analytic tools might not understand or have access to the underlying algorithms used by the system, resulting in users taking inappropriate actions. Boire (2015) also warned of users of predictive analytical tools placing too much faith on outputs, stressing the importance of understanding the accuracy limitations of predictive models. Many large firms are actively developing predictive analytic tools to address issues faced by users and customers (Schoenherr & Speier-Pero, 2015).

Halladay (2013) and Zliobaite et al. (2012) stated that large data processing organizations, including IBM, KXEN, Oracle, Portrait Software, and SAS Institute, are leading the development of predictive analytical tools for organizational use. However, Borkar et al. (2012) stated open source projects are leading the way in advancing data analytic tool development, though no open-source parallel database offering is available on the market. The distinction between big data analytics and traditional data analytics is rapidly disappearing (Boyd & Crawford, 2012). As data sets generated by organizations continue to grow, Boyd and Crawford (2012) stated the term *big data* includes large data sets, as well as the computational power, applied to data sets to gain insights into human networking and community.

Many types of data exist for analysis, but log file data is the most common type of data available to IT staff and the most informative for IT staff investigating user complaints (Oliner, Ganapathi, & Xu 2012). Two types of log file data are typically available for data analysis: event log messages and state log messages (Nagaraj et al., 2012). Nagaraj et al. (2012) stated the data contained within event log messages describe a condition at some point in time, while state log messages contain the value of a system variable. An event log message might state that a server crashed and rebooted, while a state log message reported that a hard drive is 95% full (Nagaraj et al., 2012). According to Oliner et al. (2012), log files can contain both event log messages and state log messages the same alert, such as a server crashed at time 23:15, and the previous shutdown occurred 23 days ago. The end goal of a predictive analytic tool applied to log files is to perform predictive and descriptive modeling on log file data to provide actionable information to the system administrator or other IT staff (Oliner et al., 2012).

An example of predictive modeling is a system monitoring the number of bad login attempts (Abdullah, Pillai, & Cai, 2015). According to Abdullah et al. (2015), when the number of invalid login attempts is above the statistical average, the system can generate a message, notifying IT staff of a potential issue with an authentication system, or of a security incident such as a denial of service attack is in progress. An example of descriptive modeling is a system detecting high disk usage on servers, but also receiving event codes for software installation and notes the high disk usage is occurring outside of business hours, so the event is ignored without IT staff receiving a notification unless the issue persists after the installation is complete and deletion of temporary files occurs (Kumar & Sharma, 2015). By learning how to perform tasks without needing to alert IT staff, the predictive analytic tool prevents IT workers from becoming distracted from irrelevant information (Kumar & Sharma, 2015).

According to Budgagaa et al. (2016), many challenges exist when implementing data analytic tools. One challenge is the role of historical information in data sets (Budgagaa et al., 2016). IT staff must determine the relevance of historical information on detecting and preventing future incidents. Domingos (2012) stated that too much irrelevant data results in the inability of the data analytic tool to quickly return relevant results. IT staff must also consider how to code data input into the system (Shazmeen et al. (2013). Shazmeen et al. (2013) demonstrated effective data coding could improve the performance of predictive models. Another challenge is configuring the system to provide statically significant results without returning random data or overwhelming the technician with irrelevant information (Blossom, 2014). Without the thoughtful implementation of a data analytic tool, the possibility may not exist to gain insights into IT incidents (Frizzo-Barker & Chow-White, 2014).

Another challenge arising during the implementation of a data analytic tool, such as Hadoop, relates to security breaches (Alabi et al., 2015). System logs and help desk tickets can contain proprietary, sensitive, or confidential data (Rasheed, 2014). Alabi et al. (2015) noted data analytic systems could be a source of deliberate or accidental data leakage. Due to information protection laws and policies, IT staff could be prohibited from analyzing certain stored data, therefore facing barriers to implementing data warehouse analytic measures on certain systems used by contractors and employees (Rasheed, 2014).

Burdon and Harpur (2014) stated that as users become aware of the use of predictive analytic tools they could alter their actions, preventing the tool from generating useful predictions. A possibility exists that system administrators would incorrectly classify a change to prevent the data analytic tool from reporting a potential issue that would result in the system administrator having to perform additional work (Burdon & Harpur, 2014). Burdon and Harpur (2014) stated the importance of IT staff understanding the benefits from implementing predictive analytic tools, and that management should not use the predictive analytic tools to assign blame, especially in highly distributed teams. IT employees should implement predictive analytic tools to improve service offerings (Halladay, 2013).

Obtaining funding to implement and support data analytic tools is another challenge faced by IT staff (Amarasingham et al., 2014). The cost of implementing the data analytic tool must outweigh the benefits of spending funds elsewhere in the organization (Richey et al., 2016). The use of open-source tools, such as Hadoop, reduces the cost of implementing data analytic systems within organizations, though trained experts are necessary to implement and support an instance of Hadoop (Qureshi, 2014). Even with increases in computing power each year, processing massive data sets requires significant computational power (Borkar et al., 2012). Organizations can leverage cloud computing to process large data sets to determine if any value is discernable before purchasing large computing systems capable of quickly performing data analytics (Olden, 2011).

Transition and Summary

The objective of this quantitative, quasi-experimental study is to examine whether IT staff using organizational knowledge generated from data warehouse analytical measures reduces the number of IT incidents reported by global users of IT in an international pharmaceutical company. Section 1 of the study contains the foundation of the study which includes: (a) the background of the study, (b) the problem and purpose statement, (c) the nature of the study, (d) the research question, (e) the hypotheses, (f) the theoretical framework, (g) the definition of terms, (h) the assumptions, (i) limitations and delimitations, (j) the significance of the study, and (k) a review of the professional literature related to the problem statement.

Grounded in organizational learning theory, the focus of this study is the use of predictive analytics to reduce the number of issues reported by users of IT applications in a global pharmaceutical company. Section 2 contains additional information on the research method and design, the population and sample, ethical research, data collection, data analysis techniques, and reliability and validity. Included in Section 3 will be a presentation of the findings, applications professional practice, implications for social change, recommendations for action, recommendations for further studies, reflections, and a summary of study conclusions.

Section 2: The Project

Section 2 contains expanded information related to the topics I presented in Section 1. This section includes discussions of the (a) purpose of the study, (b) role of the researcher, (c) participants, (d) method and research design, (e) population and sampling, (f) concerns relating to ethical research, (g) data collection instrument, (h) data collection and organization techniques, and (i) reliability and validity.

Purpose Statement

The purpose of this quantitative, quasi-experimental study was to examine whether IT staff use of organizational knowledge generated from data warehouse analytical measures reduced the number of IT incidents reported by global users of IT within an international pharmaceutical company headquartered in Germany. There were three independent variables in this study. The first independent variable was application system name; I analyzed 14 application systems in this study, seven as a control group, and seven as a treatment group. The second independent variable was Time 1, which consisted of the number of support calls received during an initial 30-day period before applying data warehouse analytical measures to the treatment. The application of the treatment, data warehouse analytical measures, was the third independent variable. The dependent variable was Time 2, consisting of the number of support calls received for all applications during a 30-day period when the treatment group had data analytical measures applied. This study's implications for positive social change included potential contributions of new knowledge and insights which may lead to financial savings for leaders within the pharmaceutical industry, and the ability of employees to spend more time developing innovative therapies benefiting patients.

Role of the Researcher

Researchers conducting quantitative studies must maintain objectivity and eliminate sources of bias and researcher influence when interpreting data and interacting with participants (Venkatesh, Brown, & Bala, 2013). As the sole researcher for this quantitative study, I was responsible for ethically collecting data and examining relationships between research variables (see Frels & Onwuegbuzie, 2013). I was also responsible for organizing, storing, evaluating, interpreting, and ensuring the integrity of data to test the study hypotheses and answer the research questions. Maintaining the accuracy and privacy of employee data gathered for this study was of critical importance (see Myers & Venable, 2014).

I have experience providing various types of IT support within large enterprise environments, working with ticketing systems, and interacting with help desk and general IT staff, though I have not previously researched the use of predictive analytics within organizations. Data in this study came from a help desk ticketing system and not directly from users. I obtained permission from the Walden University Institutional Review Board (IRB approval number 10-21-16-0440891) prior to gathering research data. By working with archival data from a ticketing system, I avoided potential ethical issues by not forming direct relationships with participants. Given that I could see confidential and personal data, I ensured all confidential and personal remained private (see Ferdowsian, 2011).

As specified in the Belmont Report (1979) protocol, I ensured that participants in this study were protected under the three fundamental ethical principles: (a) respect of persons, (b) beneficence, and (c) justice. I completed the National Institutes of Health training regarding the protection of human participants, and took the necessary safeguards to protect participants' rights during research (see Appendix A). I obtained permission from the human resources business partner for IT within the company participating in this study to use data from the help desk ticketing system, and to implement the treatment.

Participants

My goal in this study was to examine whether IT staff can leverage existing data from multiple sources to discover patterns in service interruptions and prevent the occurrence of future outages. I conducted this study in a global pharmaceutical company with approximately 45,000 employees, headquartered in Germany. I secured approval from the company's local head of IT prior to collecting data (see Appendix B).

Because data regarding service interruptions came from a help desk ticketing system, I had no direct contact with users and there were no direct participants. Users contacting the global help desk included internal employees, contractors, interns, and external partners who were all are over the age of 18 and are not considered vulnerable individuals. Although there were no participants directly participating in the study, I obtained permission from the Institutional Review Board (IRB approval number 10-21-16-0440891) at Walden University and complied with all policies regarding the confidentiality and protection of data acquired from the help desk tool (see Damianakis &

Woodford, 2012; Silberman & Kahn, 2011). I followed the ethical guidelines provided by Walden University throughout the study.

Data from the help desk system can include a caller's name, username, phone number, email address, and reported problems (Ma et al., 2011). I de-identified the data after extracting the data from the help desk ticketing system to maintain user and company anonymity (see Johnson, 2014). I stored all data in a locked fireproof safe on an encrypted drive, and will securely delete all extracted data after 5 years.

Research Method and Design

In this study, I worked to examine of the relationships between the use of predictive analytics and the number of reported IT incidents. The research method was quantitative, and the design was quasi-experimental. The business problem and the research question were both factors influencing my selection of the research method and design.

Method

The methodology of this study was quantitative. Researchers choose quantitative designs to examine if a significant relationship (or relationships) exists between two or more research variables (Applebaum, 2012; Goertz & Mahoney, 2012). I determined that a quantitative method was most appropriate for this study, based on the examples of Boronico et al. (2011) and Järveläinen (2013) who used quantitative designs to examine the relationships between IT support call resolution costs and incident prevention.

A qualitative method was not appropriate for this study because I was interested in understanding the relationship between variables, and not the lived experiences, actions, or motivations of users. Many researchers have used qualitative methods to understand and examine perceptions of IT managers regarding the management of operations (Oun & Bach, 2014). For instance, Baskerville et al. (2014) used a qualitative method to investigate the balance between IT incident prevention and response, while Rivard and Lapointe (2012) used a qualitative method to understand how people implementing IT systems respond to user resistance. Though qualitative studies are helpful to researchers wanting to understand how the use of data analytics improves organizational performance, the qualitative method alone would not have been useful in answering my research question.

According to Fassinger and Morrow (2013), mixed-methods studies use a combination of qualitative and quantitative methods; therefore, mixed-method could have been an appropriate method for this study. Haun et al. (2015) conducted a mixed-methods study to understand how to improve IT services offered to veterans by enhancing the synchronization, integration, and standardization of Veterans Affairs patient-facing systems. Eng et al. (2014) also conducted a mixed-methods study to understand how offering an online knowledge management system impacts health care provider outcomes. Due to time constraints in gathering data and challenges obtaining objective data from users, I did not employ this method in the study.

Research Design

In this study, my goal was to test the relationships between three independent variables (a) application (application system name), (b) Time 1 (the number of reported IT incidents during the first 30-day time period), (c) treatment (the application of the treatment), and one dependent variable, Time 2 (the number of reported IT incidents during the second 30-day time period). The design of this study was quasi-experimental, a design used by researchers performing mixed, between-within subjects ANOVA, also known as a split-plot ANOVA or SPANOVA (Edmonds & Kennedy, 2012). This design was appropriate because of the availability of archival data from a help desk ticketing system. Sheikh et al. (2013) proposed using a quasi-experimental design to understand how best to deploy limited resources while examining the relationship between cost-effectiveness and healthcare IT. Schäfferling and Wagner (2013) used a SPANOVA to investigate if investors recognized IT as a strategic organizational asset. Caine and Hanania (2013) conducted a SPANOVA to understand how patient privacy preference changed with the implementation of electronic health records.

Alternative research designs that I considered included descriptive research, correlational research, and experimental research (Gravetter & Forzano, 2015). Descriptive research is most suitable for researchers seeking to describe the current status of variables (Edmonds & Kennedy, 2012), which was not applicable to the research question of this study. Researchers choose correlational designs to understand the scope of variable relationships when investigators are not manipulating variables in an environment, and could be appropriate for scholars performing future research on IT processes (Breen, Holm, & Karlson, 2014). Kim and Park (2012) used a correlational design to verify an extended technological acceptance model using health information technology. Researchers employ true experimental designs when establishing a cause and effect relationship among variables (Cooper & Schindler, 2013). In this study, it was not possible for me to conduct a true experimental design by randomly assigning employees within the company to application groups. Subsequently, I examined a naturally occurring process of calls made to a help desk in real time, in a natural environment. In addition, my focus was on the number of calls to the help desk; thus, the population under study was comprised of those calls, not individual participants.

Population and Sampling

The number of issues reported to help desks vary depending on systems used, organizational culture, user preferences and skill, and help desk structure (Kim, Lee, & Joshi, 2013). In large organizations, employees can use many types of applications and systems to process increasing volumes and varieties of data (Fan & Bifet, 2013). Some applications might be used by all employees, such as a time clock reporting system or an email application. Specialized application systems, such as a program for processing executive compensation, may only have a few users. In this study, the population and sample were not application users, but rather the calls received by a help desk, with each call linked to applications and systems in use within the selected organization.

The population for this study included all calls logged and linked to application systems registered in the configuration database. The configuration database was the authoritative repository for all application systems in use within the organization. For a user to open a help desk ticket, there must be a configuration item (e.g., system name) within the configuration database to which to assign the ticket. The population of all calls linked to registered application systems aligned to my research question asking if using organizational knowledge generated from data warehouse analytical measures reduces the number of reported IT incidents.

There are multiple methods that I could have used for selecting the sample application systems, including probabilistic and non-probabilistic sampling typologies (Bryman, 2012). I chose purposive sampling, defined by Bryman (2012) as a selective or nonprobabilistic sampling technique, to select applications commonly supported by infrastructure application support teams within the organization. I conducted this study with support from the leaders of the infrastructure tower within IT who, in turn, benefited by learning the answer to the research question which they could use to guide decisions on optimizing IT processes and resource usage.

According to Suri (2011), multiple purposeful sampling strategies exist, and each is a valid technique, though researchers select the type best fitting the purpose of the study. Advantages of purposive sampling include: (a) researchers' ability to collect rich data, (b) greater study validity, and (c) access to large samples of people (Barratt et al., 2015; Suri, 2011). However, there are some disadvantages to purposive sampling, including: (a) access to experts who can provide information on the population and sample, (b) bias towards existing beliefs, (c) lack of population representativeness, (d) high costs, (e) generalizations, (f) confirmatory bias, and (g) hidden bias (Barratt, Ferris, & Lenton, 2015; Suri, 2011).

Multiple methods existed for choosing the application set to select for analysis in this study, including: (a) applications with the highest call volume, (b) applications supporting the most critical business functions, (c) applications with the most users, (d) management preference, and (e) applications with the most reported issues (Gravetter & Forzano, 2015). Because the purpose of the study was to learn whether using organizational knowledge generated from data warehouse analytical measures reduces the number of reported IT incidents, the underlying assumption was that IT leadership would like to reduce the number of reported IT incidents, and I selected application systems with the highest call volume under the control of infrastructure management.

After selecting the sampling method, researchers need to determine the appropriate sample size required to interpret the correlational strength between research variables (Field, 2013). An adequate sample size is necessary to ensure the reliability of the research findings (Wallen & Fraenkel, 2013). I conducted an a priori power analysis for a repeated measures, within-between interaction ANOVA using G*Power 3.1.9.2 to determine the appropriate sample size for this study. G*Power is a computer program used to performance statistical power analyses when performing tests for correlation and regression analyses (Faul, Erdfelder, Buchner, & Lang, 2009).

Researchers need to calculate the: (a) effect size, (b) alpha value, (c) number of groups, and (d) number of measures to calculate the sample size when performing quantitative research and using tests such as a SPANOVA (Muijs, 2011). The use of a medium effect size ($f = .15$) is an appropriate effect size for this proposed study. Cohen (1992) stated that the actual medium effect size is .1304. The medium effect size used in G*Power is .15, which is approximately 13.07% greater than Cohen's (1992) medium effect size. After conducting an a priori power analysis with a medium effect size ($f = .15$), $\alpha = .05$, G*Power indicated a minimum sample size of 210 is necessary to achieve a power of .80 (see Appendix C). Increasing the sample size to 406 will increase power to .99 (see Figure 1). I selected 14 application environments for this study, selecting those with a minimum of 29 support calls during the first sampling period to ensure a minimal sample size of 406 service desk calls.

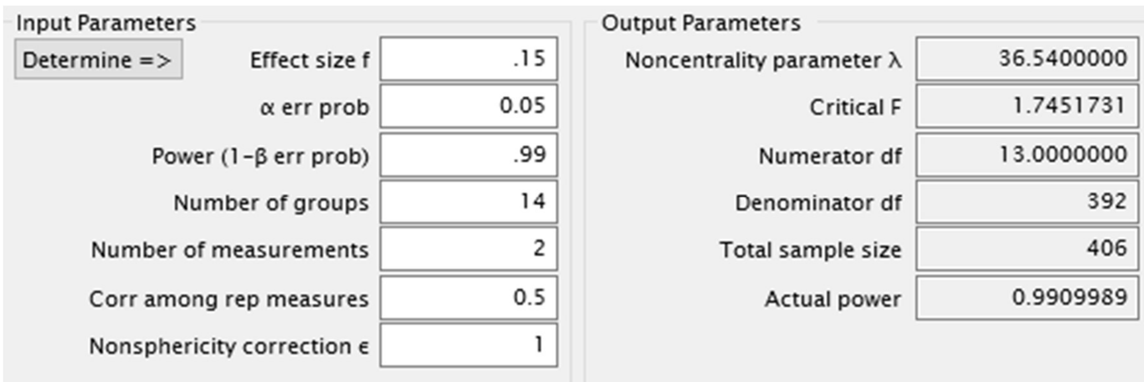


Figure 1. G*Power sample size program output.

Ethical Research

Researchers must follow ethical guidelines to protect research participants, including: (a) voluntary participant participation, (b) participant understanding of study purpose and procedures, (c) participant privacy, (d) avoidance of bias, (e) secure storage of research data, and (f) preventing conflicts of interest between researcher and participant (Terrell, 2012). Data for this study consisted of records regarding the number of incidents reported to a global help desk for selected applications. As there were no participants directly involved with this study, there was no consenting process for individuals, no process for participants to withdraw from the study, and no incentives. Despite the lack of individual participants, I obtained permission from the local Head of IT within the company participating in this study to use archival data from the help desk ticketing system, and to implement the treatment (see Appendix B). The local Head of IT could rescind permission to access service desk call information at any time

After obtaining permission to begin collecting data, I gathered data from the organization's global help desk system. Each incident reported to the help desk could contain personally identifiable information. I did not collect this information, only the total number of incidents reported for each selected application. I will maintain digital and physical records for 5 years in a locked filing cabinet that I have sole access. Furthermore, I encrypted all digital records to protect against inadvertent disclosure. To protect the identity of the company, from this point forward, I will use the pseudonym *PharmaCo* (e.g., Tilley & Woodthorpe, 2011) for the organization under study. All employee names will remain anonymous to protect individual personnel.

Data Collection

The data collection process is a critical component of quantitative research (Terrell, 2012). Details regarding data collection are important to ensure replication of the study and to increase awareness regarding data accuracy and validity (Terrell, 2012). In this section is a discussion of the study instrumentation for collecting and storing research data, information on the data collection technique, and strategies to ensure validity.

Instrumentation

The variables chosen for this study were important components in exploring the relationships between the number of reported IT incidents and data warehouse analytic measures. The research variables in this study were: (a) application (application system name), (b) Time 1 (support calls received during Time 1), (c) treatment (application of the treatment), and (d) Time 2 (support calls during Time 2).

The first independent variable was application (application system name). There were 14 application systems analyzed in this study, seven as a control group, and seven receiving the treatment. Application was a nominal variable. The second independent variable was Time 1, which consisted of the number of support calls received during an initial 30-day period before applying data warehouse analytical measures to the treatment. Time 1 was a ratio variable. Treatment (application of the treatment or data warehouse analytical measures) was the third independent variable. The treatment group variable was a dichotomous variable, with 0 representing the treatment was not applied, and 1 representing application of the treatment. The dependent variable was Time 2, consisting of the number of support calls received for all applications during a 30-day period where the treatment group had data analytical measures applied. Time 2 was a ratio variable.

I recorded the number of reported incident during Time 1 and Time 2. A reported incident occurred when a user contacted a global service desk representative to report any issue impacting the ability to use the application as intended. Users could report incidents via telephone, email, and instant message. When a user contacted a help desk representative, the help desk agent created a help desk ticket, and requested basic information including the user name, contact information, the application not operating as expected, and basic troubleshooting questions. The help desk agent could perform basic troubleshooting steps with the user or forward the ticket to a team responsible for the application. Once a help desk agent closes a ticket, the agent closing the ticket entered the resolution, whether the resolver found any fault with the system, and verified the ticket contains the correct application causing the fault or issue reported by the caller. All data logged in the help desk system was exportable, allowing IT staff to perform data analysis on all calls made by users.

I collected data from the help desk ticketing system and record all tallies for the variables in Microsoft Excel 2013, an application released on January 29, 2103 (see Figure 2). Microsoft Excel 2013 is an application used by researchers to store structured data, and the most commonly used predictive analytic tool (Greco & Aiss, 2015; Halladay, 2013; Valverde et al., 2014). Data in the Excel spreadsheet was input into IBM SPSS Statistics Version 23. SPSS version 23, released on March 3, 2015, is a statistical software package used by researchers to process data, analyze trends, and test assumptions to arrive at accurate conclusions (García-Morales et al., 2013; Halladay, 2013; Schäfferling & Wagner, 2013; Schoenherr & Speier-Pero, 2015).

application	time1	time2	treatment
Application 1			0
Application 2			0
Application 3			0
Application 4			0
Application 5			0
Application 6			0
Application 7			0
Application 8			1
Application 9			1
Application 10			1
Application 11			1
Application 12			1
Application 13			1
Application 14			1

Figure 2. Example of the instrument for collecting research variables.

The 14 application environments selected had the most support calls between July 1, 2015 and July 1, 2016. I randomly assigned the top 14 application environments to application treatment Group 1 or application treatment Group 2. Alternative methods of choosing applications included: (a) those deemed most critical to running the business, (b) applications with the highest support costs, and (c) applications with the most users. At the end of both Time 1 and Time 2, I recorded the number of calls received by the service desk for each application. Each call stored under a unique identifier or ticket counted as one call, and added to the appropriate variable in the instrument.

I used the randomize function built into Microsoft Excel 2013 to randomize the assignment of applications to the treatment and nontreatment groups. I listed each application system name in Column A, and generated a random number in Column B using the command *=rand()*. I sorted Column B from largest to smallest, expanding the selection when prompted by Microsoft Excel 2013 to include Column A. The first seven applications did not receive the treatment, and the second set of seven applications received the treatment.

I did not capture the content of support calls, as IT managers were interested in reducing the total number of support calls and the content of each call is not relevant to this study. I made available the total number of calls made for each application (see Appendix D). To determine the number of incidents reported to the global service desk, I entered the service desk tool and searched for all calls received under the application name. By searching for all calls made under each application system environment, I ensured construct validity, as help desk staff linked each call to the impacted application system environment. The possibility did not exist to review every call received by the help desk to verify help desk representatives linked each ticket to the proper application system environment.

I ran the export of calls for each application system twice and verified identical counts of calls received for each respective application system to ensure interrater reliability. Each call logged under a unique call number counted as a single call to ensure internal consistency. I exported all calls for a specified date range and filter the results by application system name to maintain consistent results (test-retest reliability). Anyone with access to the help desk ticketing system would obtain identical results by running the same search.

Data Collection Technique

Before initiating the research, I obtained permission from the local head IT (see Appendix B) to collect information from the help desk ticketing system for the 14 applications relevant to the study. To determine the top 14 applications in use at the company under the control of the infrastructure tower within IT, I contacted the employee responsible for the ticketing system, who has agreed to generate a report. Information on the total number of help desk calls per system was available to a limited number of employees. Each of the 14 applications had a unique system name that can be input into the help desk ticketing system.

To access the ticketing system, I connected to the company network and open Internet Explorer (IE). Once in IE, I typed in the name of the ticketing system, which opened the interface for VMware Service Manager, version 9.1.9 (see Figure 3). For each application, I searched for the number of calls during Time 1, and again during Time 2. To search for the number of tickets, I clicked the search button (see Figure 4), the select (a) *Open*, (b) *Open – Resolved*, (c) *Open – Unresolved*, (d) *Closed*, and (e) *Closed –Resolved*. I then entered the date range corresponding first to Time 1 (*Logged Date From* to *Logged Date To*), and again for Time 2. The final field to populate was *Configuration Item*, which is the unique system name (see Figure 5). After entering all of the information, I clicked on the *Search* button, which caused the system to provide a report of all calls opened during the specified period for the specific system. I recorded the number of calls for each system in a Microsoft Excel 2013 spreadsheet, as described in the Instrumentation section (see Appendix E).

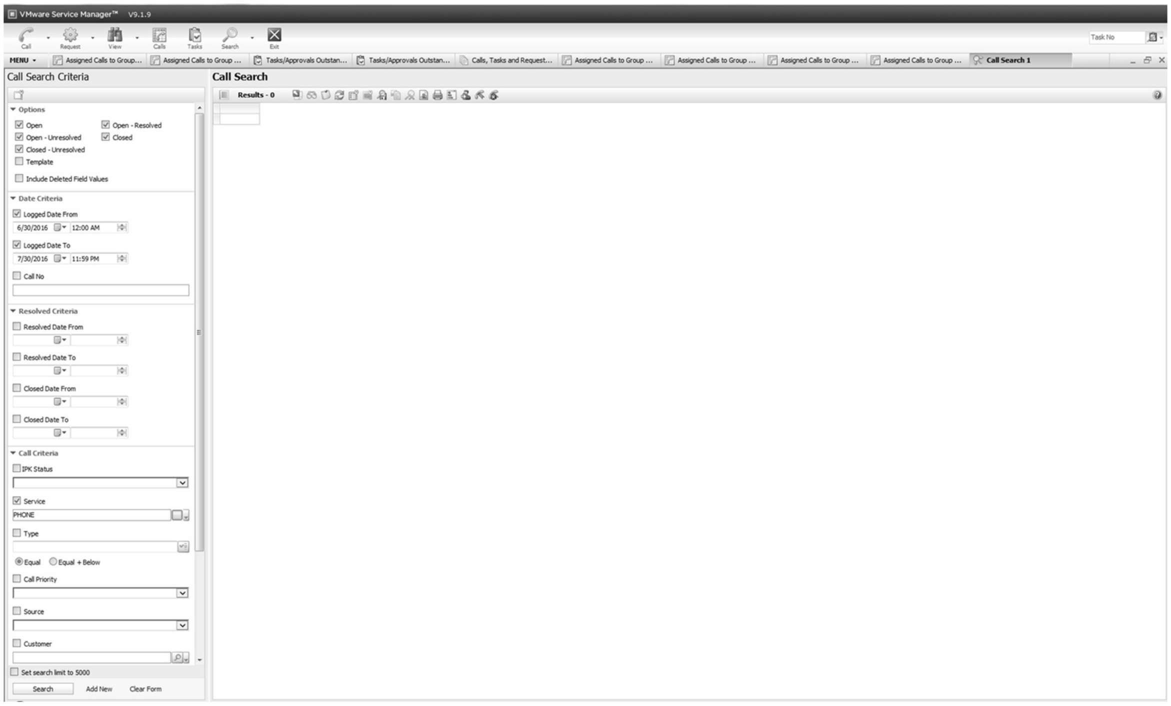


Figure 3. Search page of the help desk ticketing system.

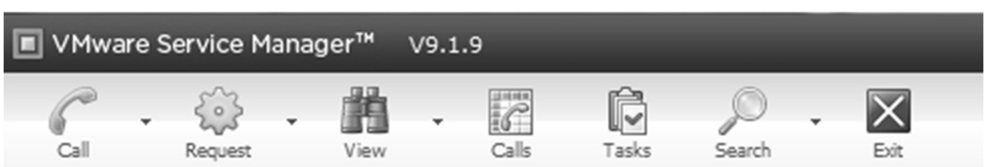


Figure 4. Interface for launching searches.

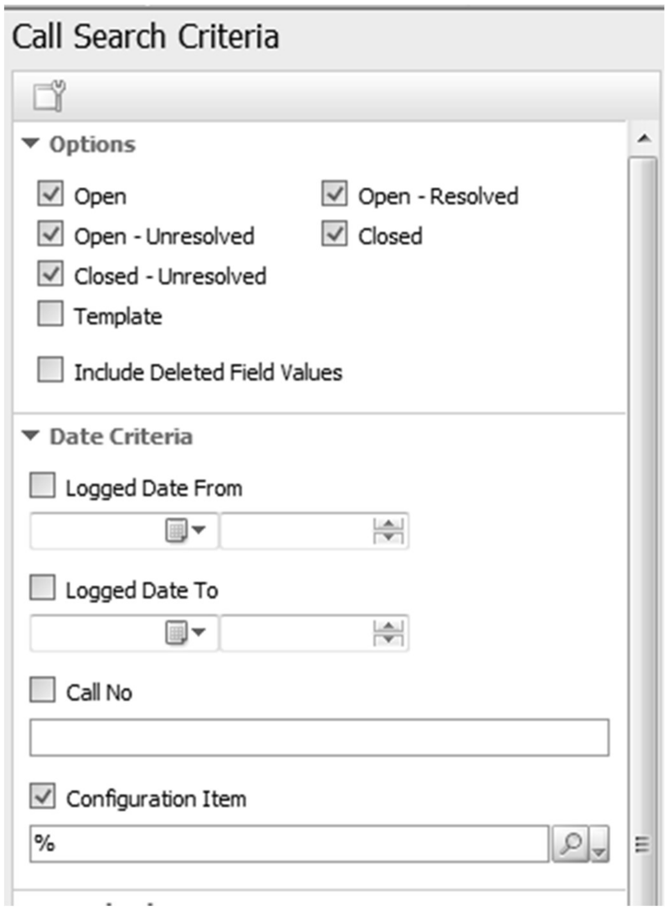


Figure 5. Interface for inputting search criteria.

The advantage of using the service desk tool for obtaining the number of calls was that the service desk tool is the official repository for all reported IT incidents. By using the official repository, I had reliable data and did not rely on opinions about whether or not users are experiencing greater or fewer incidents. The disadvantage of using the official ticketing system for obtaining the total number of IT incidents was that information to analyze is only available about issues users experience with IT systems and report to the help desk.

Data Analysis Technique

My overarching research question was: Does using organizational knowledge generated from data warehouse analytical measures reduce the number of reported IT incidents at an international pharmaceutical company headquartered in Germany? My specific research question was: Does the use of data warehouse analytic measures in a treatment group create a statistically significant reduction in the number of reported IT incidents among seven applications over a 30-day period, when compared to a group of seven additional applications used as a control group? To answer the specific research question, I used four hypotheses to examine the relationship between the independent research variables of (a) application name, (b) support calls during Time 1, and (c) treatment group; and the dependent variable, support calls during Time 2.

The first independent variable was application, which was the application system name, a nominal variable. The second independent variable was Time 1, which consisted of the number of support calls received during an initial 30-day period before applying data warehouse analytical measures to the treatment. Time 1 was a ratio variable. The treatment group (application of the treatment or data warehouse analytical measures) was the third independent variable. The treatment group variable was a dichotomous variable, with 0 representing the treatment was not applied, and 1 representing the application of the treatment. The dependent variable was Time 2, consisting of the number of support calls received for all applications during a 30-day period where the treatment group had data analytical measures applied. Time 2 was a ratio variable. The treatment group variable had a 0 representing no application of the treatment, and a one representing application of the treatment. No other coding was necessary, as the number of incidents reported during Time 1 and Time 2 were nominal.

A low likelihood existed that I would encounter a situation where data was unavailable on the number of calls received for applications during Time 1 or Time 2. If data were unavailable for any application during Time 1, I would exclude that application and select the next application in the help desk ticketing system meeting the criteria described in the research design section. If data were unavailable during Time 2, I would exclude the application from the analysis. As data were pulled from the help desk ticketing system, there was a low probability that the number of reported incidents would be unavailable.

I recorded the number of incidents received for all 14 applications during Time 1 and Time 2 in Microsoft Excel 2013 and imported the data into IBM SPSS Statistics Version 23 to perform data analysis to test each hypothesis and ultimately answer the overarching research question. The hypotheses were:

$H1_0$: There is no significant difference in the number of incidents reported during Time 1 between treatment and nontreatment applications.

$H1_a$: There is a significant difference in the number of incidents reported during Time 1 between treatment and nontreatment applications.

$H2_0$: There is no significant difference in the number of incidents reported during Time 2 between treatment and nontreatment applications.

$H2_a$: There is a significant difference in the number of incidents reported during Time 2 between treatment and nontreatment applications.

$H3_0$: There is no significant difference in the number of incidents reported in the treatment application group during Time 1 and Time 2.

$H3_a$: There is a significant difference in the number of incidents reported in the treatment application group during Time 1 and Time 2.

$H4_0$: There is no significant difference in the number of incidents reported in the non-treatment application group during Time 1 and Time 2.

$H4_a$: There is a significant difference in the number of incidents reported in the non-treatment application group during Time 1 and Time 2.

The statistical test selected for this study was between-within subjects ANOVA, also known as a split-plot ANOVA or SPANOVA. SPANOVA tests are appropriate to use with quasi-experimental designs comparing three or more variables (Caine & Hanania, 2013; Edmonds & Kennedy, 2012; Schäfferling & Wagner, 2013; Sheikh et al., 2013). SPANOVA inferential statistical tests combine between-subjects designs and within-subject designs (Pallant, 2013; Tabachnick & Fidell, 2013). Alternative non-parametric and parametric statistical techniques considered for this study but not appropriate due to the nature of the question and scale of measurement of the research variables defined included: (a) Chi-square for goodness of fit, (b) Chi-square for independence, (c) McNemar's Test, (d) Cochran's Q Test, (e) Kappa Measure of Agreement, (f) Mann-Whitney U Test, (g) Wilcoxon Signed Rank Test, (h) Kruskal-Wallis Test, (i) Friedman Test, (j) independent-samples *t* test, (k) Paired-samples *t* test, (l) one-way between-groups ANOVA, (m) one-way repeated-measures ANOVA, (n) two-way analysis of variance (between groups), (o) multivariate analysis of variance (MANOVA), and (p) analysis of covariance.

The between-subjects design was useful for analyzing the relationship between the number of calls received for applications that receive the treatment and those that do not ($H1_0$ and $H2_0$). The within-subjects design was suitable for analyzing the relationship between the number of support calls received for all applications during Time 1 and Time 2 ($H3_0$ and $H4_0$). Researchers can use IBM SPSS to combine between-subjects and within-subjects variables into one analysis (Pallant, 2013). I chose to use IBM SPSS for the data analysis of this study due to my familiarity with the computer application. In addition to the SPANOVA test, I used IBM SPSS to report the descriptive statistics for the research variables, including: (a) means, (b) standard deviations, and (c) ranges.

Pallant (2013) stated six general assumptions apply to the use of the SPANOVA parametric technique. The first assumption was the variable measured, the dependent variable, was an interval or ratio (Pallant, 2013). In this study, the dependent variable, Time 2, was a ratio variable. The second assumption was the use of random sampling (Pallant, 2013). As previously described, I randomly assigned applications to the treatment group, through the selection of the 14 applications was not random, and comprised the 14 applications systems with the most support calls between July 1, 2015 and July 1, 2016. The third assumption was that observations would be independent of each other (Pallant, 2013). Each call to the service desk was an independent event, and users had no knowledge of the applications included in the study (application), or which applications will belong to the treatment group (treatment). The fourth assumption was that the sample selected from the population follows a normal distribution (Pallant, 2013). The fifth assumption was that the sample selected from the population has equal variances, or homogeneity of variance (Pallant, 2013). The final assumption was that the pattern of intercorrelations among the levels of the within-subjects variables (Time 1 and Time 2) should be identical (Pallant, 2013).

When researchers perform quantitative research, and use tests, such as a SPANOVA, the necessity exists to calculate the effect size, alpha value, number of groups, and number of measures to calculate the sample size (Muijs, 2011). The use of a medium effect size ($f = .15$) was an appropriate effect size for this proposed study. Cohen (1992) stated that the actual medium effect size is .1304. The medium effect size used in G*Power is .15, which is approximately 13.07% greater than Cohen's (1992) medium effect size.

I conducted an a priori power analysis for a repeated measures, within-between interaction ANOVA using G*Power 3.1.9.2 to determine the appropriate sample size for this study. G*Power is a computer program used to performance statistical power analyses when performing a test for correlation and regression analyses (Faul, Erdfelder, Buchner, & Lang, 2009). After conducting an a priori power analysis with a medium effect size ($f = .15$), $\alpha = .05$, G*Power indicated a minimum sample size of 210 is necessary to achieve a power of .80 (see Appendix C). Increasing the sample size to 406 will increase power to .99. I selected 14 application environments for this study, selecting those with a minimum of 29 support calls during the first sampling period to ensure a minimal sample size of 406 service desk calls.

After performing the SPANOVA test, I interpreted the results from IBM SPSS. I determined if the Levene's test of equality of error variances, which tests for the homogeneity of variance, was not violated (significance value that is larger than .05). If the significance value was less than or equal to .05, then the assumption is variances were not equal, which could indicate the populations sampled were not identical.

The second verification was Box's *M* test of equality of covariance matrices, to determine if the pattern of intercorrelations among the levels of the within-subjects variable are identical (significance value that is larger than .001). If $p < .001$, then the possibility existed the dependent variables do not follow a multivariate normal distribution, and there are unequal sample sizes.

The third analysis was the interaction effect, to determine if there is a change in the number of incidents reported to the service desk over time between the two application groups. A significance level greater than the chosen alpha level of .05 from the Wilks' lambda distribution means the interaction effect was not statistically significant. A statistically significant result indicates an interaction effect between the independent variables, leading to a careful interpretation of the main effects (Pallant, 2013).

After determining the statistical significance of the interaction effect, I evaluated the main effects of the independent variables (application, Time 1, and treatment). Any Wilks' lambda distribution with a *p* value less than .05 meant that there was a statistically significant effect on that variable. If the *p* value was less than or equal to the alpha ($p \leq$ 05), then I would reject a null hypothesis, and the result was statistically significant. If the *p* value was greater than alpha ($p > .05$), then I would fail to reject a null hypothesis, and the result was not statistically significant. The final analysis was of the effect size of the result. The guidelines proposed by Cohen (1992) were .02 indicated a small effect, .15 indicated a moderate effect, and .35 indicated a large effect for multiple and partial correlation tests.

Study Validity

Internal and external validity are two categories of a validity of concern to researchers performing quantitative research (Whitley & Kite, 2013). Researchers performing quasi-experimental research with strong internal validity feel confident about concluding that any change observed in the dependent variable was the result of the treatment, and not the result of some other factor (Rovai, Baker, & Ponton, 2013). External validity represents the extent to which researchers' can apply their findings to other settings (Campbell & Stanley, 1966). Multiple threats to internal and external validity exist that scholars must control for while performing research (Baskerville et al., 2014; Bryman, 2012; Campbell & Stanley, 1966; Rivard & Lapointe, 2012; Simon & Goes, 2013).

Threats to internal validity included procedures, treatments, or experiences of the participants involved in the study can or may threaten the ability of researchers to draw conclusions about cause and effect (Bryman, 2012; Simon & Goes, 2013). Campbell and Stanley (1966) identified eight variables that can be a threat to internal validity: (a) history, (b) maturation, (c) testing, (d) instrumentation, (e) statistical regression, (f) selection, (g) experimental mortality, and (h) selection interactions. To protect against instrumentation and selection interaction threats to internal validity, users who contacted the help desk to report issues with applications were unaware that any of the applications they use were part of this or any other study. IT staff were also unaware of the applications selected for analysis to guard against selection bias. The setting for this study could not change, as the research question is specific to the organizational type. To control for the history and maturation threats to validity, I selected the 30-day periods to measure the research variables on the day I received URR approval to begin research. I made a note of any changes to the underlying IT landscape and organization that could jeopardize internal validity. There was no testing, and applications selected for the study will not have decommissioning dates during the year of the study. The threat of statistical regression was relevant and a possibility existed that the applications with the most calls would see the greatest reduction in calls after applying the treatment. The use of a control group mitigated the statistical regression threat to internal validity (Faul et al., 2009; Gravetter & Forzano, 2015; Motulsky, 2015).

Bryman (2012) stated threats to external validity involve characteristics of the sample, setting, or timing that threatens the ability of researchers to generalize the research findings to other populations. Threats to external validity included: (a) interactions of treatment and testing, (b) interactions of treatment and selection, (c) reactive effects of experimental arrangements, and (d) multiple treatment interference (Bryman, 2012). The design of this study was an *equivalent time samples design*, a design type that typically does not face threats to external validity, though investigators performing research with as equivalent time samples design still face potential threats from the four types of external validity (Campbell & Stanley, 1966).

There was be no pretesting, which reduces the first threat to external validity (Campbell & Stanley, 1966). To guard against selection bias, I selected the top 14 applications with the most calls from users to the global help desk. I randomly assigned the applications to treatment and non-treatment groups to control for any differences in the treatment groups from unconscious bias (Cooper & Schindler, 2013). The environment and experimental arrangement for this study could not change, as the research question is specific to the organization type. To protect against the threat of multiple treatment interference, I did not apply multiple treatments to the same application.

Threats to statistical conclusion validity threaten researchers' ability to draw valid statistical inferences, including: (a) reliability of the instrument, (b) data assumptions, and (c) sample size (García-Pérez, 2012). According to García-Pérez (2012), researchers can face two types of error when performing research. Type I error involves researchers finding a correlation where none exists, while type II error occurs when researchers find that no correlation exists when in reality one does (García-Pérez, 2012; Gravetter & Forzano, 2015). The instrument for this study was a Microsoft Excel 2013 spreadsheet that I used to capture each call made to the help desk. An incident was a call made to the global help desk that a help desk representative recorded under a unique identification number in a ticket. Once an IT worker closed a ticket for an incident, a future call from the same user counted as an additional incident.

Transition and Summary

In Section 2, I reintroduced the purpose of this study and provided additional details on the research method and design. I also presented the population and sample for the study, information on performing ethical research, data collection instrumentation and technique, data analysis technique, and study validity. In Section 3, I will present: (a) an overview of the study, (b) a presentation of the findings, (c) the application to professional practice, (d) the implications for social change, (e) the recommendations for action, (f) the recommendations for future study, (g) reflections, and (h) a summary and conclusion of the study.

Section 3: Application to Professional Practice and Implications for Change

Section 3 contains the results of the analysis presented in Section 2. This section includes (a) an overview of the study, (b) a presentation of the findings, (c) a discussion of applications to professional practice, (d) a discussion of implications for social change, (e) recommendations for actions, (f) recommendations for further study, and (g) reflections. I close this chapter with a summary and study conclusions.

Overview of Study

The purpose of this quantitative, quasi-experimental study was to examine whether IT staff use of organizational knowledge generated from data warehouse analytical measures reduced the number of IT incidents reported by global users of IT within an international pharmaceutical company headquartered in Germany. Because of limited resources, IT workers often need to balance the time and money spent resolving issues against time spent proactively preventing issues (Boronico, Zirkler, & Siegel, 2011). My goal in this study was to fill a gap in the scholarly and professional understanding of how IT managers should allocate limited IT resources within an organization, specifically in the area of incident prevention and incident resolution, through the use of data warehouse analytical measures.

My overarching research question was: Does using organizational knowledge generated from data warehouse analytical measures reduce the number of reported IT incidents at an international pharmaceutical company headquartered in Germany? My specific research question was: Does the use of data warehouse analytic measures (treatment) in a treatment group (Time 2) create a statistically significant reduction in the number of reported IT incidents among applications (application) over a 30-day period, when compared to a group of seven additional applications used as a control group (Time 1)? The associated hypotheses were:

$H1_0$: There is no significant difference in the number of incidents reported during Time 1 between treatment and nontreatment applications.

$H1_a$: There is a significant difference in the number of incidents reported during Time 1 between treatment and nontreatment applications.

$H2_0$: There is no significant difference in the number of incidents reported during Time 2 between treatment and nontreatment applications.

$H2_a$: There is a significant difference in the number of incidents reported during Time 2 between treatment and nontreatment applications.

$H3_0$: There is no significant difference in the number of incidents reported in the treatment application group during Time 1 and Time 2.

$H3_a$: There is a significant difference in the number of incidents reported in the treatment application group during Time 1 and Time 2.

$H4_0$: There is no significant difference in the number of incidents reported in the nontreatment application group during Time 1 and Time 2.

$H4_a$: There is a significant difference in the number of incidents reported in the nontreatment application group during Time 1 and Time 2.

There was not a significant difference in the number of reported incidents during Time 1, between treatment and nontreatment applications ($H1_0$). While there was a difference in the number of reported incidents during Time 2 between treatment and nontreatment applications ($H2_0$), the difference was not significant. This indicates that warehouse analytic measures may indeed reduce the number of calls, when compared to the nontreatment group; however, this reduction was not statistically significant. There was a difference in the number of incidents reported in the treatment application group during Time 1 and Time 2 ($H3_0$), and a difference in the number of incidents reported in the nontreatment application group during Time 1 and Time 2 ($H4_0$), though none of these differences were statistically significant. Based on these results, I found no statistically significant differences between the number of reported IT incidents during Time 1 and Time 2 among treatment and nontreatment applications, and I failed to reject all four null hypotheses.

Presentation of the Findings

This study included four variables. The first independent variable was application, which was the application system name. I analyzed 14 application systems in this study: seven as a control group, and seven as a treatment group. The second independent variable was Time 1, which consisted of the number of support calls received during an initial 30-day period before applying data warehouse analytical measures to the treatment. The third independent variable was treatment, which consisted of the application of data warehouse analytical measures. The dependent variable was Time 2, consisting of the number of support calls received for all applications during a 30-day period when the treatment group had data analytical measures applied.

The four research variables were related to the study hypotheses. The first hypotheses included all calls received for all applications during Time 1. The second hypotheses included all calls received for all applications during Time 2. The third hypotheses included all calls received for applications receiving the treatment during Time 1 and Time 2. The fourth hypotheses included all calls received for applications that did not receive the treatment during Time 1 and Time 2.

To analyze the relationship between the variables, I performed a mixed, between-within subjects ANOVA, also known as a split-plot ANOVA or SPANOVA (see Edmonds & Kennedy, 2012; Pallant, 2013; Tabachnick & Fidell, 2013). I made the decision to use a quasi-experimental design because of the availability of archival data from a help desk ticketing system, and the use of this design by researchers examining similar relationships (e.g., Sheikh et al., 2013). The design and statistical tests were critical to determining the significance of using data warehouse analytic techniques to reduce reported IT incidents.

Descriptive Statistics

The data for this study were generated from a help desk ticketing system, as described in Section 2, and captured using the instrument in Appendix E. The instrument consisted of an Excel 2013 spreadsheet with a column for each of the four variables: (a) application (application system name), (b) Time 1 (support calls received during Time 1), (c) treatment (application of the treatment), and (d) Time 2 (support calls during Time 2). No data were missing or unavailable during data collection because I created the dataset from existing systems. Therefore, I did not have to make decisions regarding how to address missing data. I imported the data from the instrument created in Excel 2013 into IBM SPSS Version 23, and used SPSS to perform statistical analysis on the data collected from the help desk ticketing system. IBM SPSS is a statistical software package used by researchers to process data, analyze trends, and test assumptions (García-Morales et al., 2013; Halladay, 2013; Schäfferling & Wagner, 2013; Schoenherr & Speier-Pero, 2015).

I used IBM SPSS to report the descriptive statistics for the research variables, including: (a) means, (b) standard deviations, and (c) ranges. During Time 1, the help desk reported 21,512 calls for the 14 IT systems analyzed, with the number of calls for each application ranging from 440 to 5,597 (M = 1,537; SD = 1,417). Of the calls received during Time 1, 9,111 were from applications that were not in the treatment group (M = 1,302; SD = 891), and 12,401 were from applications that were in the treatment group (M = 1,772; SD = 1,852).

During Time 2, the help desk reported 20,759 calls for the 14 IT systems analyzed, with the number of calls for each application ranging from 184 to 5,691 (M = 1,483; SD = 1,495); this reduction of 753 calls (3.5%) was not significant. Of the calls received during Time 2, 8,408 were from applications that were not in the treatment group (M = 1,201; SD = 837), this reduction of 703 calls (7.7%) was not significant, and 12,351 were from applications that were in the treatment group (M = 1,764; SD = 1,989), a not significant reduction of 50 calls (4%). Table 2 contains the descriptive statistics for the research variables.

Table 2

Descriptive Statistics

	Nontreatment group	Treatment group	Treatment and nontreatment group
Reported incidents during Time 1	9,111	12,401	21,512
Reported incidents during Time 2	8,408	12,351	20,759
Mean of reported incidents during Time 1	1,302	1,772	1,537
Standard deviation of reported incidents during Time 1	891	1,852	1,417
Mean of reported incidents during Time 2	1,201	1,764	1,483
Standard deviation of reported incidents during Time 2	837	1,989	1,495

Inferential Statistical Analysis

I performed a mixed, between-within subjects ANOVA to examine the relationship between the number of help desk calls received for 14 applications, two time periods, and the application of a treatment. The first independent variable was application. The between-subjects test variable was the application of the treatment (treatment). The within-subjects variable was the nontreatment applications (Time 1). The dependent variable was Time 2.

General assumptions apply to parametric techniques such as mixed between-within subjects ANOVA, as listed in Section 2 (see Tabachnick & Fidell, 2013). The first assumption to test was that the sample selected from the population had equal variances, or homogeneity of variance (see Pallant, 2013). To test for homogeneity of variance, I performed Levene's test. Researchers can use a Levene's test to determine if equal variances exist between two or more groups (Field, 2013). Field (2013) stated that Levene's test is not significant when the significance value calculated by the test is greater than .05 ($p > .05$). The significance value for Time 1 is .150, and .078 for Time 2 (see Table 3). If the significance value was less than or equal to .05, then the assumption is variances were not equal, which could indicate that the populations sampled were not identical. Because both significance values are greater than .05, the significance value is not significant, indicating that no violation of homoscedasticity occurred.

Table 3

Test of Homogeneity of Variance

Variable	F	*df*1	*df*2	Sig.
Time 1	2.368	1	12	.150
Time 2	3.710	1	12	.078

The second parametric test I performed was Box's test of equality to determine whether two or more covariance matrices are equal. A significance value greater than .001 indicates no violation of the assumption (Pallant, 2013). The calculated significance value was .042, which was greater than .001 (see Table 4).

Table 4

Box's Test of Equality of Covariance Matrices

Box's M	F	*df*1	*df*2	Sig
10.017	2.736	3	25920.000	.042

If the significance value was less than .001, then it would have been possible that the dependent variable (Time 2) did not follow a multivariate normal distribution, and there were unequal sample sizes (see Pallant, 2013). The calculated significance value of .042 indicates that the dependent variable (Time 2) followed a normal distribution with equal sample sizes.

The final assumption to test was for the pattern of intercorrelations among the levels of the within-subjects variables (Time 1 and Time 2), which should be identical (see Pallant, 2013). The interaction effect is not statistically significant if the significance level for the Wilk's Lambda is greater than the selected alpha level of .05 (Pallant, 2013). The calculated Wilks' Lambda significance value was .407 (see Table 5). According to Pallant (2013) a statistically significant result could indicate an interaction effect between the independent variables (application, Time 1, and treatment). A calculated significance value of .407 indicates there was no interaction between the independent variables (application, Time 1, and treatment).

Table 5

Treatment Group Wilks' Lambda Multivariate Test

Value	F	Hypothesis df	Error df	Sig.	Partial eta squared
.942	.737	1.000	12.000	.407	.058

After determining that the interaction effect was not significant, my next step was to assess the main effects for each of the independent variables. The value for Wilks' Lambda for treatment and non-treatment applications within Time 1 and Time 2 was .924, with a significance value of p = .407, partial eta squared = .76, with treatment and nontreatment groups showing a reduction in the number of reported incidents across both time periods. Because the p value was greater than .05, I concluded that there is not a statistically significant effect for treatment. The lack of a statistically significant effect for treatment indicates that there was not a significant change in the number of reported incidents across the two different time periods among the applications receiving the treatment and the applications that did not. The main effect for time was insignificant (p = .407), and I failed to reject $H1_0$ (there is no significant difference in the number of incidents reported in the treatment application group during Time 1 and Time 2) and $H2_0$ (there is no significant difference in the number of incidents reported during Time 2 between treatment and nontreatment applications).

After testing the within-subjects variables, my next test was the between-subjects interaction. The value for Wilks' Lambda for the nontreatment applications within Time 1 and Time 2 was .924, F (1, 12) = .980, with a significance value of p = .342, partial eta squared = .76, with the nontreatment group showing a nonsignificant reduction in the number of reported incidents across both time periods (see Table 6). Because the p value was greater than .05, I concluded that there is not a statistically significant effect for time. A not a statistically significant effect for time suggests that there was not a significant change in the number of reported incidents across the two different time periods. The main effect comparing the application of treatment and time was not significant, $p = .528$, partial eta squared = .034, suggesting no difference from the application of the treatment (see Table 7). The main effect for time was insignificant ($p = .528$), and I failed to reject $H3_0$ (there is no significant difference in the number of incidents reported in the treatment application group during Time 1 and Time 2.) and $H4_0$ (there is no significant difference in the number of incidents reported in the nontreatment application group during Time 1 and Time 2).

Table 6

Time Wilks' Lambda Multivariate Test

Value	F	Hypothesis df	Error df	Sig.	Partial eta squared
.942	.980	1.000	12.000	.342	.076

Table 7

Tests of Between-Subjects Effects

F	df	Error df	Sig.	Partial eta squared
.423	1	12	.528	.034

The final analysis was of the effect size of the result. The guidelines proposed by Cohen (1992) were .02 indicated a small effect, .15 indicated a moderate effect, and .35 indicated a large effect for multiple and partial correlation tests. The partial eta squared for all applications during Time 1 and Time 2 was .058, and .076 for nontreatment applications, indicating a small effect for time. The partial eta squared for applications receiving treatment was .034, indicating a small effect for treatment. Time appears to have had a greater effect on the number of reported incidents to the help desk than the application of the treatment, though no effect was statistically significant.

My overarching research question was: Does using organizational knowledge generated from data warehouse analytical measures reduce the number of reported IT incidents at an international pharmaceutical company headquartered in Germany? The specific research question was: Does the use of data warehouse analytic measures (treatment) in a treatment group (Time 2) create a statistically significant reduction in the number of reported IT incidents among applications (application) over a 30-day period, when compared to a group of seven additional applications used as a control group (Time 1)? Based on the analysis of the data, the use of data warehouse analytic measures did not significantly reduce the number of reported IT incidents at an international pharmaceutical company headquartered in Germany.

Analysis of Findings

The theoretical framework used in this study was organizational learning theory, a theory researchers use to explain how members of organizations learn from previous mistakes to arrive at a different outcome (see Dodgson, 1993). As applied to this study, IT managers at an international pharmaceutical company used deutero-learning to reflect on how organizational members discover new insights to solve problems. IT employees deployed a Hadoop cluster capable of processing server log data to learn if the possibility existed to prevent future issues from occurring by leveraging existing organizational data, going beyond simple problem analysis associated with single-loop learning (Baskerville et al., 2014).

Organizational learning theory is one theory to explain how IT workers could reduce the number of reported IT incidents by analyzing existing data. Prior to implementing the treatment, IT workers did not routinely examine log file data to prevent system interruptions from impacting users of technology. The analysis of log data was a novel way for employees to learn from previous mistakes to arrive at a different outcome in the form of a reduction of reported incidents to a global help desk. The possibility exists that employees examining log data discovered methods for optimizing system availability without using data analytic techniques.

My findings correlated with existing research grounded in the literature regarding the use of using organizational knowledge to reduce IT incidents. Songsangyos et al. (2012) noted IT workers could use expert systems to reduce the workload of help desk staff, but failed to demonstrate an actual reduction in workload. Datta and Lade (2015) were unable to demonstrate a reduction in support calls by moving help desk employees to specialized roles within the organization. Kaur and Kinger (2013) posited that monitoring services could detect problems and automatically alert the responsible team to take corrective action without human intervention, but did not associate the implementation of advanced systems with a reduction in IT incidents. Similar to findings by prior researchers searching for methods of reducing IT service interruptions, the use of data analytic techniques did not lead to a significant reduction in the number of reported IT incidents.

Applications to Professional Practice

Data warehouse analytic tools consist of powerful machine learning and statistical techniques that business leaders can leverage to gain insights to predict and improve business processes (Boire, 2015; Halladay, 2013; Moses & Chan, 2014; Qureshi, 2014). While many researchers demonstrated how machine learning algorithms are cost-effective alternatives to hiring staff for manual programming and data mining activities, I did not find a significant reduction in the number of reported IT incidents within an international pharmaceutical company that applied data analytic techniques to server log data (Domingos, 2012; Parikh, Kakad, & Bates, 2016). Leaders from similar organizations need to decide if resources should be spent on developing systems to apply data analytic techniques to log data to reduce reported IT incidents, or spend resources on alternative process improvement methods, such as ITIL, Six Sigma, and total system intervention for system failure. ITIL, Six Sigma, and total system intervention for system failure were all demonstrated as methods for improving and optimizing organizational processes (Li et al., 2011; Nakamura & Kijima, 2011; Talla & Valverde, 2013; Tehrani & Mohamed, 2011; Valverde, Saadé, & Talla, 2014).

Employees within organizations can use predictive analytic tools to discover patterns in data, enabling informed decision making (Jinka et al., 2012). If IT workers wish to discover patterns in server log data otherwise hidden by the huge volume of information available, predictive analytic tools could be beneficial (Blossom, 2014; Zliobaite et al., 2012). While predictive analytic tools can help employees discover patterns in data, the findings of this study do not support the premise that the use of predictive analytic tools can lead to a significant reduction in reported IT incidents.

Implications for Social Change

A program put in place to enhance the availability and integrity of IT systems containing data relevant to patient safety ultimately benefits patients and increases profits (Armando et al., 2015). Managers within organizations should note that employee job satisfaction could increase when end-users of technology do not face issues, and IT workers can feel more job satisfaction when time is spent developing technical solutions instead of resolving issues (Siti-Nabiha et al., 2012). End-users could be less resistant to adopt new technologies when the technology works as expected, allowing for the rapid and successful rollout of IT systems (Rivard & Lapointe, 2012).

The implications for positive social change included contributions of new knowledge, insights, and financial savings to leaders within the pharmaceutical industry. IT leadership within the studied organization gained new knowledge and are now aware that the use of predictive analytic tools alone did not lead to a reduction in reported IT incidents. Resources should be spent on investigating alternative methods for reducing IT incidents so that employees can focus on developing innovative therapies benefiting patients. Newly gained insights include the realization that a possibility exists to save money by discontinuing investments in data warehouse analytic tools installed with the intention of reducing reported IT incidents.

Society may benefit from this research, as programmers can develop improved data analytic tools that can support reducing the number of incidents users of technology experience. Business and government agencies can support initiatives that improve the experiences of technology users, so users are more productive and feel greater satisfaction in their lives. By spending limited resources on tools that improve the experiences of users, organizational leaders can create a more desirable workplace for employees, leading to gains in efficiency and employee fulfillment.

Recommendations for Action

Predictive analytic tools have a wide range of uses, including: (a) banking, (b) social media, (c) college recruitment and retention, (d) crime prevention, (e) insurance modeling, (f) warranty claims processing, (g) talent acquisition, (h) e-commerce, and (i) supply chain management (Burdon & Harpur, 2014; Greco & Aiss, 2015; Jinka et al., 2012; Moses & Chan, 2014; Schoenherr & Speier-Pero, 2015). The findings from this study were that a significant reduction in reported IT incidents did not occur from the use of data warehouse analytic tools on server log data for seven applications within an international pharmaceutical company. While users employing predictive analytic tools can discover patterns in data otherwise hidden in huge volumes of information, the implementation of predictive analytic tools alone does not guarantee that identified patterns can improve a business process (Blossom, 2014; Zliobaite et al., 2012).

Based on the findings of this study, I recommend IT professionals examine existing data warehouse analytic tools to determine if the possibility exists to optimize the software, with a focus on error detection and prevention. Software developers should note that out of the box, currently available data analytic tools are not effective at predicting future IT system failures. As developers release improved versions of data analytic tools, organizational leaders can replicate the mixed between-within subjects ANOVA used in this study to determine if the use of updated tools can significantly reduce the number of calls received be help desk employees. Information regarding user experiences with data analytic tools could be shared at technical conferences, trade shows, and industry publications.

Recommendations for Further Study

Many tools exist that are capable of processing massive datasets. Hadoop®, Apache Spark™, and Storm™, are generic tools that users can leverage to process a wide variety of data sets. While these software programs are powerful, developers did not design these open source products specifically for analyzing server log data, nor were the programs designed to reduce reported IT incidents. Future researchers should determine if coupling Hadoop with additional software such as Hive can improve the predictive analytic capabilities of the system. IT staff could also investigate the use and value of propriety software specifically designed to detect and prevent IT incidents, such as Splunk, or open source programs such as Graylog.

Future researchers should also investigate if predictive analytic tools generate actionable data based on application or system type. The possibility exists that a tool would have a higher likelihood of predicting a hardware failure than predicting a user will forget a password. Organizational leaders would need to determine the value of implementing a system that only reduces certain types of IT incidents.

Another area for future research is regarding the data contained in system log files. The possibility exists that log data within the studied organization was not detailed enough, and that options exist to configure applications to produce verbose logs which contain more information for an analytic tool to process. Organizational leaders would need to determine if the cost of collecting additional system log data, coupled with the potential security risks of collecting and storing detailed log data, is worth a potential reduction in IT incidents.

The data in this study only applied to one international pharmaceutical company's users and applications. Due to the thousands of application systems in use within the company, no possibility existed to examine all of the application systems used by employees, nor did a possibility exist to test out every tool capable of performing data analytics. With over 45,000 users of IT, I could not follow-up on every support ticket opened by users within the organization. The possibility exists that many user calls were not due to system faults, but users requesting assistance using an application or access to an application. Future researchers could investigate the content of reported IT incident help desk tickets to learn if the incident was the fault of an IT system. IT managers could investigate if implementing tools to suggest what steps IT staff could take to prevent an outage would benefit an organization, along with tools that could predict the impact of actions taken by IT staff.

Reflections

My desire for completing a quantitative study analyzing the use of software to reduce IT incidents arose from discussions on basing decisions on fact instead of intuition or opinion. Organizational leaders can use dashboards to report on metrics over a period of time, such as the number of help desk calls received, the number of overdue help desk tickets, and the time to resolution for help desk calls, but without a statistical analysis of the data, conclusions could be misleading. This study was the first step in determining the value of using data warehouse analytic techniques within an organization to improve IT service offerings for end users.

The most challenging hurdle I encountered was in finding scholarly sources related to a cutting edge technology. Many published sources of information were from conference proceedings, industry trade journals, and vendor case studies. My hope is that researchers will perform additional studies on the topic of using big data to improve common sources of friction within organizations and submit their research for peer review.

Based on the findings from this study, I concluded that many variables exist relating to the successful deployment of data analytic platforms such as Hadoop. This study was the first step is determining a way to use existing organizational data to improve IT service offerings. During my research I discovered the implications for successfully data mining large data sets within IT goes well beyond reducing reported IT service interruptions. Additional uses of data warehouse analytic techniques within IT include IT security, user training, application system sizing, and compliance monitoring.

Summary and Study Conclusions

IT leadership in many organizations must determine the best use of limited resources to support business functions. While users expect IT staff to quickly resolve reported IT incidents, users would benefit if the possibility existed to prevent incidents from occurring. One possible way to reduce IT support calls is for IT employees to use of data warehouse analytic measures on system log data to determine if IT workers can implement preventative measures before an issue impacts users.

Based on the analysis of the data from an international pharmaceutical company headquartered in Germany, the use of data warehouse analytic measures did not significantly reduce the number of reported IT incidents. An analysis of these findings by IT leadership might be beneficial for planning how to use limited resources to support business functions, and for developing the next generation of data analytic tools to support the identification of error trends within application systems.

References

Abdullah, A. B., Pillai, T. R., & Cai, L. Z. (2015). Intrusion detection forecasting using time series for improving cyber defence. *International Journal of Intelligent Systems and Applications in Engineering*, *3*(1), 28-33. http://dx.doi.org/10.18201/ijisae.83441

Ackerman, M. S., Dachtera, J., Pipek, V., & Wulf, V. (2013). Sharing knowledge and expertise: The CSCW view of knowledge management. *Computer Supported Cooperative Work*, *22*, 531-573. http://dx.doi.org/10.1007/s10606-013-9192-8

Akinnuwesi, B. A., Enikuomehin, O. A., Uzoka, F. M. E., Onwudike, O. C., Osamiluyi, A., & Aribisala, B. S. (2014). Electronic helpdesk support system in tertiary institutions in developing countries. *International Journal of Computer and Information Technology, 3*, 1280-1291. Retrieved from http://www.ijcit.com

Aggarwal, B., Bhagwan, R., De Carli, L., Padmanabhan, V., & Puttaswamy, K. (2011). Deja vu: fingerprinting network problems. In *Proceedings of the Seventh Conference on Emerging Networking Experiments and Technologies*, 28-40. http://dx.doi.org/10.1145/2079296.2079324

Alabi, O., Beckman, J., Dark, M., & Springer, J. (2015). Toward a data spillage prevention process in Hadoop using data provenance. In *Proceedings of the 2015 Workshop on Changing Landscapes in HPC Security*, 9-13. http://dx.doi.org/10.1145/2752499.2752502

Allman, E. (2012). Managing technical debt. *Communications of the ACM*, *55*(5), 50-55. http://dx.doi.org/10.1145/2160718.2160733

Amarasingham, R., Patzer, R. E., Huesch, M., Nguyen, N. Q., & Xie, B. (2014). Implementing electronic health care predictive analytics: considerations and challenges. *Health Affairs*, *33*, 1148-1154. http://dx.doi.org/10.1377/hlthaff.2014.0352

Applebaum, M. (2012). Phenomenological psychological research as science. *Journal of Phenomenological Psychology*, *43*(1), 36-72. http://dx.doi.org/10.1163/156916212X632952

Argyris, C. (1994). Good communication that blocks learning. *Harvard Business Review*, *72*(4), 77-85. http://dx.doi.org/10.1225/5386

Argyris, C., & Schön, D. (1978). *Organizational Learning: A theory of action perspective*. Reading, MA: Addison-Wesley

Armando, A., Bezzi, M., Metoui, N., & Sabetta, A. (2015). Risk-based privacy-aware information disclosure. *International Journal of Secure Software Engineering*, *6*(2), 70-89. http://dx.doi.org/10.4018/IJSSE.2015040104

Armstrong, S., Simer, L., & Spaniol, L. (2011). Models of technology management at the community college: The role of the chief information officer. *New Directions for Community Colleges*, *2011*(154), 87-95. http://dx.doi.org/10.1002/cc.449

Ashcraft, P. L., Cummings, S. M., Fogle, B. G., & Valdez, C. D. (2015). *AskIT Service Desk Support Value Model* (No. LA-UR-15-26291). Los Alamos National Laboratory (LANL).

Asthana, A., & Okumoto, K. (2012). Integrative software design for reliability: Beyond models and defect prediction. *Bell Labs Technical Journal*, *17*(3), 37-59. http://dx.doi.org/10.1002/bltj.21557

Babalâc, C. C., & Uda, M. (2014). Change management essentials to manage business flexibility. *Management Research and Practice*, *6*(2), 85-91. http://dx.doi.org/10.1108/shr.2011.37210eaa.002

Barlow, J. B., Warkentin, M., Ormond, D., & Dennis, A. R. (2013). Don't make excuses! Discouraging neutralization to reduce IT policy violation. *Computers & Security*, *39*, 145-159. http://dx.doi.org/10.1016/j.cose.2013.05.006

Barratt, M. J., Ferris, J. A., & Lenton, S. (2015). Hidden populations, online purposive sampling, and external validity taking off the blindfold. *Field Methods*, *27*(1), 3-21. http://dx.doi.org/10.1177/1525822X14526838

Baskerville, R., Spagnoletti, P., & Kim, J. (2014). Incident-centered information security: Managing a strategic balance between prevention and response. *Information & Management*, *51*(1), 138-151. http://dx.doi.org/10.1016/j.im.2013.11.004

Beath, C., Becerra-Fernandez, I., Ross, J., & Short, J. (2012). Finding value in the information explosion. *MIT Sloan Management Review*, 53(4), 17-21. Retrieved from http://sloanreview.mit.edu/article/finding-value-in-the-information-explosion/

Blossom, J. (2014). The signal economy: How to target new revenues through predictive analytics. *Information Services and Use*, *34*(1-2), 17-25. http://dx.doi.org/10.3233/ISU-140724

Bober, P. (2014). Simulation for IT service desk improvement. *Quality Innovation Prosperity*, *18*(1), 47-58. http://dx.doi.org/10.12776/QIP.V18I1.343

Boire, R. (2013). Is predictive analytics for marketers really that accurate? *Journal of Marketing Analytics*, *1*(2), 118-123. http://dx.doi.org/10.1057/jma.2013.8

Borkar, V., Carey, M. J., & Li, C. (2012). Inside Big Data management: ogres, onions, or parfaits?. In *Proceedings of the 15th International Conference on Extending Database Technology*, 3-14. ACM. http://dx.doi.org/10.1145/2247596.2247598

Boronico, J. S., Zirkler, A., & Siegel, P. H. (2011). Quantifying the tradeoffs between cost and quality for systems service support. *Journal of Applied Business Research*, *12*(4), 70-88. Retrieved from http://www.cluteinstitute.com

Boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, *15*, 662-679. http://dx.doi.org/10.1080/1369118X.2012.678878

Breen, R., Holm, A., & Karlson, K. B. (2014). Correlations and Nonlinear Probability Models. *Sociological Methods & Research*, *43*, 571-605. http://dx.doi.org/10. 1177/0049124114544224

Brown, T., & Brandt, D. (2014). How high growth economies impact global information technology departments. *AI & Society*, *29*(2), 241-247. http://dx.doi.org/10. 1007/s00146-013-0478-7

Bryman, A. (2012). *Social research methods*. New York, NY: Oxford University Press.

Budgaga, W., Malensek, M., Pallickara, S., Harvey, N., Breidt, F. J., & Pallickara, S. (2016). Predictive analytics using statistical, learning, and ensemble methods to support real-time exploration of discrete event simulations. *Future Generation Computer Systems*, *56*, 360-374. http://dx.doi.org/10.1016/j.future.2015.06.013

Burdon, M., & Harpur, P. D. (2014). Re-conceptualising privacy and discrimination in an age of talent analytics. *University of New South Wales Law Journal*, *37*, 679-712. Retrieved from unsw.edu.au

Caine, K., & Hanania, R. (2013). Patients want granular privacy control over health information in electronic medical records. *Journal of the American Medical Informatics Association*, *20*(1), 7-15. http://dx.doi.org/10.1136/amiajnl-2012-001023

Campbell, D. T., & Stanley, J. C. (1966). Experimental and quasi-experimental designs for research. Chicago, IL: Rand McNally.

Chong, S., Skalka, C., & Vaughan, J. A. (2015). Self-identifying data for fair use. *Journal of Data and Information Quality*, *5*(3), 11:1-11:30. http://dx.doi.org/10.1145/2687422

Clark, C. M., Murfett, U. M., Rogers, P. S., & Ang, S. (2012). Is empathy effective for customer service? Evidence from call center interactions. *Journal of Business and Technical Communication*, 27(2) 123-153. http://dx.doi.org/10.1177/1050651912468887

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*, 155-159. http://dx.doi.org/10.1037/0033-2909.112.1.155

Cooper, D. R., & Schindler, P. S. (2013). *Business research methods* (12th ed.). New York, NY: McGraw-Hill/Irwin.

Datta, S., & Lade, P. (2015). Will this be Quick?: A case study of bug resolution times across industrial projects. In *Proceedings of the 8th India Software Engineering Conference*, 20-29. ACM. http://dx.doi.org/10.1145/2723742.2723744

Damianakis, T., & Woodford, M. R. (2012). Qualitative research with small connected communities generating new knowledge while upholding research ethics. *Qualitative Health Research*, *22*, 708-718. http://dx.doi.org/10.1177/1049732311431444

Dodgson, M. (1993). Organizational learning: a review of some literatures. *Organization Studies*, *14*, 375-394. http://dx.doi.org/10.1177/017084069301400303

Domingos, P. (2012). A few useful things to know about machine learning. *Communications of the ACM*, *55*(10), 78-87. http://dx.doi.org/10.1145/2347736.2347755

Dzuba, R. (2015). Exploring the Experiences of Call Center Employees Regarding Business Scripting (Doctoral dissertation). Retrieved from http://scholarworks.waldenu.edu/cgi/viewcontent.cgi?article=1575&context=dissertations

Edmonds, W. A., & Kennedy, T. D. (2012). *An applied reference guide to research designs: Quantitative, qualitative, and mixed methods*. Thousand Oaks, CA: Sage Publications.

Edwards, H. K. (2015). Post-implementation management in large-scale management information systems: Changes, incident reports, help desk calls, and business process. *International Journal of Business and Information*, 3(2), 207-232. http://dx.doi.org/10.1.1.472.8193

Ehrhart, K. H., Witt, L. A., Schneider, B., & Perry, S. J. (2011). Service employees give as they get: Internal service as a moderator of the service climate–service outcomes link. *Journal of Applied Psychology*, *96*, 423-431. http://dx.doi.org/ 10.1037/a0022071

Ekekwe, O. J. (2013). *Relationship between institutional frameworks and growth of SMEs in Nigeria's petroleum industry* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 1318499616)

Eng, J. J., Noonan, V. K., Townson, A. F., Higgins, C. E., Rogers, J., & Wolfe, D. L. (2014). Impact of an online medical internet site on knowledge and practice of health care providers: A mixed methods study of the spinal cord injury rehabilitation evidence project. *Journal of Medical Internet Research*, *16*(12). E26http://dx.doi.org/10.2196/jmir.3453

Fassinger, R., & Morrow, S. (2013). Toward best practices in quantitative, qualitative, and mixed-method research: A social justice perspective. *Journal for Social Action in Counseling and Psychology*, *5*(2), 69-83. Retrieved from http://psysr.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. http://dx.doi.org/10.3758/BRM.41.4.1149

Ferdowsian, H. (2011). Human and animal research guidelines: Aligning ethical constructs with new scientific developments. *Bioethics*, *25*, 472-478. http://dx.doi.org/10.1111/j.1467-8519.2011.01923.x

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

Flynn, W. C., & Philbin, G. (2014). Behind the help desk: career, salary and training expectations. *Issues in Information Systems*, *15*(2), 285-292. Retrieved from http://iacis.org

Foss, N. J., & Hallberg, N. L. (2014). How symmetrical assumptions advance strategic management research. *Strategic Management Journal*, *35*, 903-913. http://dx.doi.org/10.1002/smj.2130

Frels, R. K., & Onwuegbuzie, A. J. (2013). Administering quantitative instruments with qualitative interviews: A mixed research approach. *Journal of Counseling & Development*, 91, 184-194. http://dx.doi.org/10.1002/j.1556-6676.2013.00085.x

French, A. M. (2015). A case study on e-banking security? When security becomes too sophisticated for the user to access their information. *The Journal of Internet Banking and Commerce*, 17(2), 1-14. Retrieved from http://www.arraydev.com/commerce/jibc/

Frizzo-Barker, J., & Chow-White, P. A. (2014). Research in brief: From patients to petabytes: genomic big data, privacy, and informational risk. *Canadian Journal of Communication*, 39, 615-625. http://dx.doi.org/10.1.1.719.8768

García-Morales, V. J., Jiménez-Barrionuevo, M. M., & Gutiérrez-Gutiérrez, L. (2012). Transformational leadership influence on organizational performance through organizational learning and innovation. *Journal of Business Research*, *65*, 1040-1050. http://dx.doi.org/10.1016/j.jbusres.2011.03.005

García-Pérez, M. A. (2012). Statistical conclusion validity: some common threats and simple remedies. *Frontiers in Psychology*, 3(325), 1-11. http://dx.doi.org/10. 3389/fpsyg.2012.00325

Goertz, G., & Mahoney, J. (2012). Concepts and measurement: Ontology and epistemology. *Social Science Information*, *51*(2), 205-216. http://dx.doi.org/ 10.1177/0539018412437108

González, L. M., Giachetti, R. E., & Ramirez, G. (2005). Knowledge management-centric help desk: specification and performance evaluation. *Decision Support Systems*, *40*(2), 389-405. http://dx.doi.org/10.1016/j.dss.2004.04.013

Govindarajulu, C. (2014). A federated model for end-user computing support. *Journal of Management*, *15*(1), 71-77. http://dx.doi.org/10.4018/joeuc.2008040103

Gravetter, F. J., & Forzano, L. A. B. (2015). *Research methods for the behavioral sciences, 4th edition*. Belmont, CA: Wadsworth.

Greco, A. N., & Aiss, C. G. (2015). University presses in the twenty-first century: The potential impact of big data and predictive analytics on scholarly book marketing. *Journal of Scholarly Publishing*, *46*(2), 105-140. http://dx.doi.org/10.3138/jsp. 46.2.01

Green, A. (2016). Information overload in healthcare management: How the READ Portal is helping healthcare managers. *Journal of the Canadian Health Libraries Association*, *32*(3), 173-176. Retrieved from https://www.researchgate.net

Halladay, S. D. (2013). Using Predictive Analytics to Improve Decisionmaking. *The Journal of Equipment Lease Financing*, *31*(2), B1-B6. Retrieved from thealtagroup.com

Haun, J. N., Nazi, K. M., Chavez, M., Lind, J. D., Antinori, N., Gosline, R. M., & Martin, T. L. (2015). A participatory approach to designing and enhancing integrated health information technology systems for veterans: protocol. *JMIR Research Protocols*, *4*(1), e28. http://dx.doi.org/10.2196/resprot.3815

Hayashi, A. (2014). Thriving in a big data world. *MIT Sloan Management Review*, *55*(2), 1-8. Retrieved from http://sloanreview.mit.edu

Ibidunmoye, O., Hernández-Rodriguez, F., & Elmroth, E. (2015). Performance anomaly detection and bottleneck identification. *ACM Computing Surveys*, *48*(1), 4:1-4:45. http://dx.doi.org/10.1145/2791120

Jaiswal, V., & Levina, N. (2012). J-TRADING: Full circle outsourcing. *Journal of Information Technology Teaching Cases, 2*(2), 61-70. http://dx.doi.org/10. 1057/jittc.2012.11

Järveläinen, J. (2013). IT incidents and business impacts: Validating a framework for continuity management in information systems. *International Journal of Information Management*, *33*, 583-590. http://dx.doi.org/10.1016/ j.ijinfomgt.2013.03.001

Johnson, B. (2014). Ethical issues in shadowing research. *Qualitative Research in Organizations and Management*, *9*(1), 21-40. http://dx.doi.org/10.1108/QROM-09-2012-1099

Katzan, H. (2011). Foundations of service science management and business. *Journal of Service Science*, *1*(2), 1-16. http://dx.doi.org/10.19030/jss.v1i2.4291

Kaur, J., & Kinger, S. (2013). Analysis of different techniques used for fault tolerance. *International Journal of Computer Science and Information Technologies*, *4*, 737-741. http://dx.doi.org/10.1109/IPDPS.2011.306

Kim, J., Lee, K., & Joshi, K. (2013). Editorial Preface: Globalization of Information Technology Products and Services. *Journal of Global Information Technology Management*, *16*(3), 1-6. http://dx.doi.org/10.1080/1097198X.2013.10845639

Kim, J., & Park, H. A. (2012). Development of a health information technology acceptance model using consumers' health behavior intention. *Journal of Medical Internet Research*, *14*(5). http://dx.doi.org/10.2196/jmir.2143

Kornberger, M. (2016). The visible hand and the crowd: Analyzing organization design in distributed innovation systems. *Strategic Organization*. 2016, 1-20. http://dx.doi.org/10.1177/1476127016648499

Krogh, G., Nonaka, I., & Rechsteiner, L. (2012). Leadership in organizational knowledge creation: A review and framework. *Journal of Management Studies*, *49*, 240-277. http://dx.doi.org/10.1111/j.1467-6486.2010.00978.x

Kumar, S., & Sharma, P. (2015). Predictive analytics: Cloud computing. *International Journal*, 5(3), 6-12. Retrieved from http://www.warse.org/IJATCSE/static/pdf/file/ijatcse02532016.pdf

Kumari, S., Gupta, M. K., Khan, M. K., & Li, X. (2014). An improved timestamp-based password authentication scheme: comments, cryptanalysis, and improvement. *Security and Communication Networks*, *7*, 1921-1932. http://dx.doi.org/10.1002/sec.734. 15

Leung, N. K., & Lau, S. K. (2015). Relieving the overloaded help desk: a knowledge management approach. *Communications of the IIMA*, *6*(2), 85-96. http://dx.doi.org/10.1.1.98.1334

Li, S. H., Wu, C. C., Yen, D. C., & Lee, M. C. (2011). Improving the efficiency of IT help-desk service by Six Sigma management methodology (DMAIC) – a case study of C company. *Production Planning & Control*, *22*, 612-627. http://dx.doi.org/10.1080/09537287.2010.503321

Lund, J. R. (2012). Service desk shuffle. *The Bottom Line, 25*, 23-25. http://dx.doi.org/10.1108/08880451211229199

Ma, J., Kim, N., & Rothrock, L. (2011). Performance assessment in an interactive call center workforce simulation. *Simulation Modelling Practice and Theory*, *19*(1), 227-238. http://dx.doi.org/10.1016/j.simpat.2010.06.008

McNely, B. (2012). Big data, situated people: humane approaches to communication design. *Communication Design Quarterly Review*, *1*(1), 27-30. http://dx.doi.org/10.1145/2448917.2448923

Mithas, S., Tafti, A. R., Bardhan, I., & Goh, J. M. (2012). Information technology and firm profitability: mechanisms and empirical evidence. *MIS Quarterly*, *36*, 205-224. Retrieved from http://terpconnect.umd.edu/~smithas/papers/mithasetal2012misqprofit.pdf

Moe, N. B., Šmite, D., Hanssen, G. K., & Barney, H. (2014). From offshore outsourcing to insourcing and partnerships: four failed outsourcing attempts. *Empirical Software Engineering*, *19*, 1225-1258. http://dx.doi.org/10.1007/s10664-013-9272-x

Moses, L. B., & Chan, J. (2014). Using Big Data for Legal and Law Enforcement Decisions: Testing the New Tools. *University of New South Wales Law Journal*, *37*, 643-678. http://dx.doi.org/ 10.1016/j.poetic.2014.11.001

Motulsky, H. J. (2015). Common misconceptions about data analysis and statistics. *British Journal of Pharmacology*, *172*, 2126-2132. http://dx.doi.org/10. 1111/bph.12884

Moustaghfir, K. (2012). Explicating knowledge-based competitive advantage. *International Journal of Synergy and Research*, *1*(1), 23-38. Retrieved from http://econpapers.repec.org

Muijs, D. (2011). *Doing quantitative research in education with SPSS*. Thousand Oaks, CA: SAGE Publications Inc.

Myers, M. D., & Venable, J. R. (2014). A set of ethical principles for design science research in information systems. *Information & Management, 51*, 801-809. http://dx.doi.org/10.1016/j.im.2014.01.002

Nagaraj, K., Killian, C., & Neville, J. (2012). Structured comparative analysis of systems logs to diagnose performance problems. In *Proceedings of the 9th USENIX Conference on Networked Systems Design and Implementation*, 26-40. Retrieved from usenix.org

Nakamura, T., & Kijima, K. (2011). Total system intervention for system failures and its application to information and communication technology systems. *Systems Research and Behavioral Science*, *28*, 553-566. http://dx.doi.org/10.1002/sres.1114

National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research (1979). *The Belmont report*: *Ethical principles and guidelines for the protection of human subjects of research*. (DHEW *Publication No*. (OS) 78-0013). Washington, DC: U.S. Government Printing Office.

Nazari, K., & Emami, M. (2012). Knowledge management: from theory to practice. *Australian Journal of Business and Management Research*, *1*(11), 22-30. http://dx.doi.org/10.1002/9783527687039.ch02

Nevis, E. C., DiBella, A. J., Gould, J. M. (1995). Understanding Organizations as Learning Systems. *MIT Sloan Management Review*, *36*(2), 73-85. http://dx.doi.org/10.1016/j.infsof.2014.07.008

Noruzy, A., Dalfard, V. M., Azhdari, B., Nazari-Shirkouhi, S., & Rezazadeh, A. (2013). Relations between transformational leadership, organizational learning, knowledge management, organizational innovation, and organizational performance: an empirical investigation of manufacturing firms. *The*

International Journal of Advanced Manufacturing Technology, *64*, 1073-1085. http://dx.doi.org/10.1007/s0070-012-4038-y

Obasa, A. I., & Salim, N. (2014). Mining FAQ from forum threads: Theoretical framework. *Journal of Theoretical & Applied Information Technology*, *63*(1), 39-50. Retrieved from www.jatit.org

Oktavia, T. (2014). Implementing data warehouse as a foundation for decision support system (perspective: technical and nontechnical factors). *Journal of Theoretical and Applied Information Technology*, *60*, 476-482. Retrieved from http://www.jatit.org/volumes/Vol60No3/3Vol60No3.pdf

Olden, E. (2011). Architecting a cloud-scale identity fabric. *Computer*, *44*(3), 52-59. http://dx.doi.org/10.1109/MC.2011.60

Oliner, A., Ganapathi, A., & Xu, W. (2012). Advances and challenges in log analysis. *Communications of the ACM*, *55*(2), 55-61. http://dx.doi.org/10.1145/ 2076450.2076466

Onwuegbuzie, A. J., Leech, N. L., & Collins, K. M. T. (2012). Qualitative analysis techniques for the review of the literature. *The Qualitative Report*, *17*(28), 1-28. Retrieved from http://www.nova.edu/ssss/QR/aindex.html

Ordonez, C. (2013). Can we analyze big data inside a DBMS?. In *Proceedings of the Sixteenth International Workshop on Data Warehousing and OLAP*, 85-92. ACM. http://dx.doi.org/ 10.1145/2513190.2513198

Oun, M. A., & Bach, C. (2014). Qualitative Research Method Summary. *Journal of Multidisciplinary Engineering Science and Technology*, *1*(5), 252-258. Retrieved from http://www.jmest.org

Pallant, J. (2013). *SPSS survival manual: A step by step guide to data analysis using IBM SPSS* (5th ed). Maidenhead, England: McGraw-Hill Education.

Parikh, R. B., Kakad, M., & Bates, D. W. (2016). Integrating predictive analytics into high-value care: the dawn of precision delivery. *JAMA: The Journal of the American Medical Association*, *315*, 651-652. http://dx.doi.org/10.1001/jama.2015.19417

Provost, F., & Fawcett, T. (2013). Data science and its relationship to big data and data-driven decision making. *Big Data*, *1*(1), 51-59. http://dx.doi.org/10.1089/big.2013.150

Qureshi, B. (2014). Towards a digital ecosystem for predictive healthcare analytics. In *Proceedings of the 6th International Conference on Management of Emergent Digital EcoSystems*, 34-41. http://dx.doi.org/10.1145/2668260.2668286

Rasheed, H. (2014). Data and infrastructure security auditing in cloud computing environments. *International Journal of Information Management*, *34*(3), 364-368. http://dx.doi.org/10.1016/j.ijinfomgt.2013.11.002

Richardson, B. F., & Mahfouz, A. Y. (2014). Strategic Integration of IT and Business Service Management. *Communications of the IIMA*, *9*(2), 7. Retrieved from http://scholarworks.lib.csusb.edu

Richey, R. G., Morgan, T. R., Lindsey-Hall, K., & Adams, F. G. (2016). A global exploration of big data in the supply chain. *International Journal of Physical Distribution & Logistics Management*, *46*, 710-739. http://dx.doi.org/10.1108/IJPDLM-05-2016-0134

Rivard, S., & Lapointe, L. (2012). Information technology implementers' responses to user resistance: nature and effects. *MIS Quarterly*, *36*, 897-920. Retrieved from misq.org

Robbins, T. R., & Harrison, T. P. (2011). New project staffing for outsourced call centers with global service level agreements. *Service Science*, *3*(1), 41-66. http://dx.doi.org/10.1287/serv.3.1.41

Rovai, P. A., Baker, D. J., & Ponton, K. M. (2013). *Social science research design and statistics: A practitioner's guide to research methods and IBM SPSS analysis*. Chesapeake, VA: Watertree Press LLC.

Salah, K., & Chaudary, A. (2015). Modelling and analysis of rule-based network security middleboxes. *IET Information Security*, *9*, 305-312. http://dx.doi.org/10.1049/iet-ifs.2014.0545

Salthouse, T. A. (2011). All data collection and analysis methods have limitations: Reply to Rabbitt (2011) and Raz and Lindenberger (2011). *Psychological Bulletin, 137*, 796-799. http://dx.doi.org/10.1037/a0024843

Schäfferling, A., & Wagner, H. T. (2013). Do investors recognize information technology as a strategic asset? A longitudinal analysis of changes in ownership structure and it capability. In *European Conference on Information Systems 2013 Proceedings*, 31-43. Retrieved from http://aisel.aisnet.org

Schoenherr, T., & Speier-Pero, C. (2015). Data science, predictive analytics, and big data in supply chain management: Current state and future potential. *Journal of Business Logistics*, *36*(1), 120-132. http://dx.doi.org/10.1111/jbl.12082

Selvaraj, M. G., Jayabal, D. S., Srinivasan, T., & Balasubramanie, P. (2015). Predicting defects using information intelligence process models in the software technology project. *The Scientific World Journal*, 2015, 1-6. http://dx.doi.org/10.1155/2015/598645

Shazmeen, S. F., Baig, M. M. A., & Pawar, M. R. (2013). Performance evaluation of different data mining classification algorithm and predictive analysis. *Journal of Computer Engineering*, *10*(6), 1-6. http://dx.doi.org/0.6084/m9.figshare.1151629

Sheikh, A., Nurmatov, U. B., Cresswell, K., & Bates, D. (2013). Investigating the cost-effectiveness of health information technologies: a systematic review protocol. *BMJ Open*, *3*(12), 1-3. http://dx.doi.org/10.1136/bmjopen-2013-003737

Silberman, G., & Kahn, K. L. (2011). Burdens on research imposed by institutional review boards: The state of the evidence and its implications for regulatory reform. *Milbank Quarterly*, 89, 599-627. http://dx.doi.org/10.1111/j.1468-0009.2011.00644.x

Simon, M. K., & Goes, J. (2013). *Dissertation & scholarly research, recipes for success*. New York, NY: Dissertation Success.

Siti-Nabiha, A. K., Thum, W. Y., & Sardana, G. D. (2012). A case study of service desk's performance measurement system. *International Journal of Commerce and Management*, *22*(2), 103-118. http://dx.doi.org/10.1108/10569211211239412.

Solemon, B., Ariffin, I., Din, M. M., & Anwar, R. M. (2013). A review of the uses of crowdsourcing in higher education. *International Journal of Asian Social Science*, *3*, 2066-2073. Retrieved from http://www.aessweb.com

Somvanshi, S., Gogate, C., Sahane, A., & Kolhe, K. (2014). Online authoring tool for knowledge management system. *International Journal*, *2*(4), 205-208. Retrieved from http://ijarcsms.com

Son, G., Hazlewood, V., & Peterson, G. D. (2014). On automating XSEDE user ticket classification. In *Proceedings of the 2014 Annual Conference on Extreme Science and Engineering Discovery Environment,* 41-48. ACM. http://dx.doi.org/ 10.1145/2616498.2616549

Songsangyos, P., Niyomkha, W., & Tumthong, S. (2012). Helpdesk expert. *International Journal of Arts & Sciences,* 5(4), 137-143. Retrieved from http://www.sci.rmutsb.ac.th/

Sousa, N., Costa, C. J., & Aparicio, M. (2013). IO-SECI: A conceptual model for knowledge management. In *Proceedings of the Workshop on Open Source and Design of Communication*, 9-17. ACM. http://dx.doi.org/10.1145/ 2503848.2503850

Stevanovic, D., Vlajic, N., & An, A. (2013). Detection of malicious and non-malicious website visitors using unsupervised neural network learning. *Applied Soft Computing*, *13*, 698-708. http://dx.doi.org/10.1016/j.asoc.2012.08.028

Strohmeier, S. (2013). Employee relationship management—Realizing competitive advantage through information technology?. *Human Resource Management Review*, *23*(1), 93-104. http://dx.doi.org/10.1016/j.hrmr.2012.06.009

Suri, H. (2011). Purposeful sampling in qualitative research synthesis. *Qualitative Research Journal, 11(2),* 63-75. http://dx.doi.org/10.3316/QRJ1102063

Sykes, T. A. (2015). Support structures and their impacts on employee outcomes: A longitudinal field study of an enterprise system implementation. *MIS Quarterly*, *39,* 473-495. Retrieved from http://www.misq.org/

Tabachnick, B.G., & Fidell, L.S. (2013). *Using multivariate statistics* (6th ed). Boston, MA: Pearson Education.

Talla, M., & Valverde, R. (2013). An implementation of ITIL guidelines for IT support process in a service organization. *International Journal of Information and Electronics Engineering*, *3*, 334-341. http://dx.doi.org/10.7763/IJIEE.2013.V3.329

Tehrani, A. R. F., & Mohamed, F. Z. M. (2011). A CBR-based approach to ITIL-based service desk. *Journal of Emerging Trends in Computing and Information Sciences*, *2*, 476-484. http://dx.doi.org/10.4236/ti.2013.43022

Terrell, S. R. (2012). Mixed-methods research methodologies. *The Qualitative Report*, *17*(1), 254-280. Retrieved from http://www.nova.edu/ssss/QR/QR17-1/terrell.pdf

Tilley, L., & Woodthorpe, K. (2011). Is it the end for anonymity as we know it? A critical examination of the ethical principle of anonymity in the context of 21st century demands on the qualitative researcher. *Qualitative Research, 11*(2), 197-212. http://dx.doi.org/10.1177/1468794110394073

Valverde, R., Saadé, R. G., & Talla, M. R. (2014). ITIL-based IT service support process reengineering. *Intelligent Decision Technologies*, *8*(2), 111-130. http://dx.doi.org/10.3233/IDT-130182

Venkatesh, V., Brown, S. A., & Bala, H. (2013). Bridging the qualitative-quantitative divide: Guidelines for conducting mixed methods research in information systems. *MIS Quarterly*, *37*, 21-54. Retrieved from http://www.misq.org

Vera-Baquero, A., Colomo-Palacios, R., & Molloy, O. (2013). Business process analytics using a big data approach. *IT Professional*, *15*(6), 29-35. http://dx.doi.org/10.1109/MITP.2013.60

Vinueza, G. (2013). Saving costs using smart call routing: aligning business and IT through finance. *Journal of Information Technology Teaching Cases*, *3*(2), 96-105. http://dx.doi.org/10.1057/jittc.2013.4

Wallen, E. N., & Fraenkel, R. S. (2013). *Educational research: A guide to the process*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Wang, S., Noe, R. A., & Wang, Z. M. (2014). Motivating knowledge sharing in knowledge management systems: A quasi–field experiment. *Journal of Management*, *40*, 978-1009. http://dx.doi.org/10.1177/0149206311412192

Whitley, E. B., & Kite, E. M. (2013). *Principle of research in behavioral science.* New York, NY: Routledge.

Wu, C. W. (2016). The performance impact of social media in the chain store industry. *Journal of Business Research*. http://dx.doi.org/10.1016/j.jbusres.2016.04.130

Zliobaite, I., Bifet, A., Gaber, M., Gabrys, B., Gama, J., Minku, L., & Musial, K. (2012). Next challenges for adaptive learning systems. *ACM SIGKDD Explorations Newsletter*, *14*(1), 48-55. http://dx.doi.org/10.1145/2408736.2408746

Appendix A: Certificate of Completion – NIH PHRP

Certificate of Completion

The National Institutes of Health (NIH) Office of Extramural Research certifies that **Mark Malley** successfully completed the NIH Web-based training course "Protecting Human Research Participants".

Date of completion: 09/13/2014.

Certification Number: 1555555.

Appendix B: Permission to Collect Data

DATA USE AGREEMENT

This Data Use Agreement ("Agreement"), effective as of October 12, 2016 ("Effective Date"), is entered into by and between Mark Malley ("Data Recipient") and [redacted] "Data Provider"). The purpose of this Agreement is to provide Data Recipient with access to a Limited Data Set ("LDS") for use in scholarship/research **in accord with laws and regulations of the governing bodies associated with the Data Provider, Data Recipient, and Data Recipient's educational program.** In the case of a discrepancy among laws, the agreement shall follow whichever law is more strict.

1. Definitions. Due to the project's affiliation with Laureate, a USA-based company, unless otherwise specified in this Agreement, all capitalized terms used in this Agreement not otherwise defined have the meaning established for purposes of the USA "HIPAA Regulations" and/or "FERPA Regulations" codified in the United States Code of Federal Regulations, as amended from time to time.

2. Preparation of the LDS. Data Provider shall prepare and furnish to Data Recipient a LDS in accord with any applicable laws and regulations of the governing bodies associated with the Data Provider, Data Recipient, and Data Recipient's educational program.

3. Data Fields in the LDS. **No direct identifiers such as names may be included in the Limited Data Set (LDS).** In preparing the LDS, Data Provider shall include the **data fields specified as follows**, which are the minimum necessary to accomplish the project: access to the help desk ticketing system to obtain a list of the 14 applications with the most help desk calls, along with the number of calls received for each application from January 2015 until completion of the doctoral study.

4. Responsibilities of Data Recipient. Data Recipient agrees to:

   a. Use or disclose the LDS only as permitted by this Agreement or as required by law;

   b. Use appropriate safeguards to prevent use or disclosure of the LDS other than as permitted by this Agreement or required by law;

   c. Report to Data Provider any use or disclosure of the LDS of which it becomes aware that is not permitted by this Agreement or required by law;

   d. Require any of its subcontractors or agents that receive or have access to the LDS to agree to the same restrictions and conditions on the use and/or disclosure of the LDS that apply to Data Recipient under this Agreement; and

   e. Not use the information in the LDS to identify or contact the individuals who are data subjects.

5. Permitted Uses and Disclosures of the LDS. Data Recipient may use and/or disclose the LDS **for the present project's activities only.**

6. Term and Termination.

   a. Term. The term of this Agreement shall commence as of the Effective Date and shall continue for so long as Data Recipient retains the LDS, unless sooner terminated as set forth in this Agreement.

   b. Termination by Data Recipient. Data Recipient may terminate this agreement at any time by notifying the Data Provider and returning or destroying the LDS.

   c. Termination by Data Provider. Data Provider may terminate this agreement at any time by providing thirty (30) days prior written notice to Data Recipient.

   d. For Breach. Data Provider shall provide written notice to Data Recipient within ten (10) days of any determination that Data Recipient has breached a material term of this Agreement. Data Provider shall afford Data Recipient an opportunity to cure said alleged material breach upon mutually agreeable terms. Failure to agree on mutually agreeable terms for cure within thirty (30) days shall be grounds for the immediate termination of this Agreement by Data Provider.

   e. Effect of Termination. Sections 1, 4, 5, 6(e) and 7 of this Agreement shall survive any termination of this Agreement under subsections c or d.

7. Miscellaneous.

   a. Change in Law. The parties agree to negotiate in good faith to amend this Agreement to comport with changes in federal law that materially alter either or both parties' obligations under this Agreement. Provided however, that if the parties are unable to agree to mutually acceptable amendment(s) by the compliance date of the change in applicable law or regulations, either Party may terminate this Agreement as provided in section 6.

   b. Construction of Terms. The terms of this Agreement shall be construed to give effect to applicable federal interpretative guidance regarding the HIPAA Regulations.

   c. No Third Party Beneficiaries. Nothing in this Agreement shall confer upon any person other than the parties and their respective successors or assigns, any rights, remedies, obligations, or liabilities whatsoever.

d. Counterparts. This Agreement may be executed in one or more counterparts, each of which shall be deemed an original, but all of which together shall constitute one and the same instrument.

e. Headings. The headings and other captions in this Agreement are for convenience and reference only and shall not be used in interpreting, construing or enforcing any of the provisions of this Agreement.

IN WITNESS WHEREOF, each of the undersigned has caused this Agreement to be duly executed in its name and on its behalf.

DATA PROVIDER

Signed: [redacted]

Print Name: [redacted]

Print Title: U.S. Head of IT

DATA RECIPIENT

Signed:

Print Name: Mark Meller

Print Title: [redacted]

Letter of Cooperation from a Research Partner

[redacted]

Name: [redacted]

Title: US Head of IT

10/12/2016

Dear Mark Malley,

Based on my review of your research proposal, I give permission for you to conduct the study entitled *Proactive IT Incident Prevention: Using Data Analytics to Reduce Service Interruptions* within [redacted] As part of this study, I authorize you to access the help desk ticketing system to obtain the list of the 14 applications with the most help desk calls, along with the number of calls received for each from January 2015 until completion of the doctoral study.

We understand that our organization's responsibilities include: access to the help desk ticketing system. We reserve the right to withdraw from the study at any time if our circumstances change.

I confirm that I am authorized to approve research in this setting and that this plan complies with the organization's policies.

I understand that any identifying data collected will remain confidential and may not be provided to anyone outside of the student's supervising faculty/staff without permission from the Walden University IRB.

Sincerely,

[redacted]

Walden University policy on electronic signatures: An electronic signature is just as valid as a written signature as long as both parties have agreed to conduct the transaction electronically. Electronic signatures are regulated by the Uniform Electronic Transactions Act. Electronic signatures are only valid when the signer is either (a) the sender of the email, or (b) copied on the email containing the signed document. Legally an "electronic signature" can be the person's typed name, their email address, or any other identifying marker. Walden University staff verify any electronic signatures that do not originate from a password-protected source (i.e., an email address officially on file with Walden).

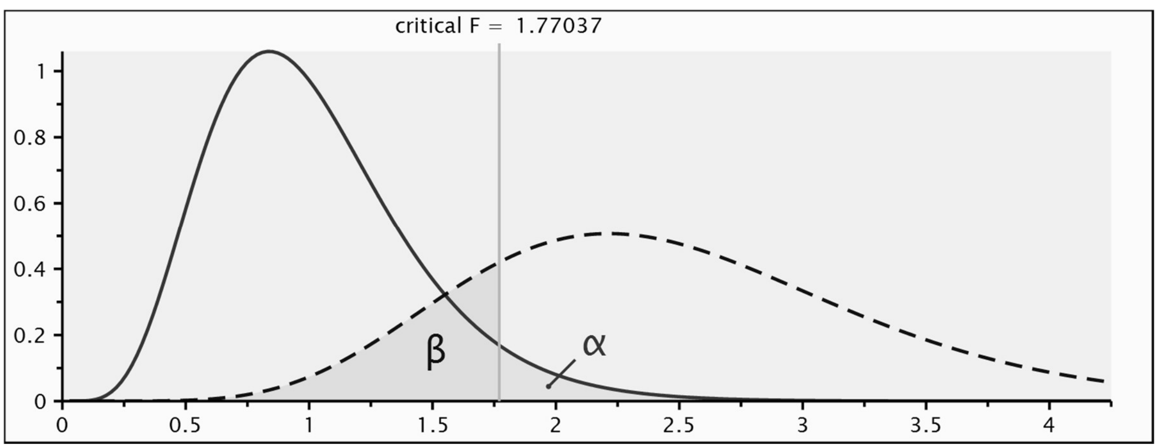
Appendix C: Protocol of Power Analyses Using G*Power 3.1.9.2

[1] -- Saturday, June 11, 2016 -- 17:41:34

F tests - ANOVA: Repeated measures, within-between interaction

Analysis: A priori: Compute required sample size

Input:	Effect size f	=	.15
	α err prob	=	0.05
	Power (1-β err prob)	=	.8
	Number of groups	=	14
	Number of measurements	=	2
	Corr among rep measures	=	0.5
	Nonsphericity correction ε	=	1
Output:	Noncentrality parameter λ	=	18.9000000
	Critical F	=	1.7703695
	Numerator d f	=	13.0000000
	Denominator df	=	196
	Total sample size	=	210
	Actual power	=	0.8007112



Appendix D: Total Number of Calls for Each Application

Application	Reported incidents during Time 1	Reported incidents during Time 2	Total reported incidents	Application of treatment
Application 1	3,194	3,012	6,206	No
Application 2	1,329	1,226	2,555	No
Application 3	1,170	1,001	2,171	No
Application 4	1,254	998	2,252	No
Application 5	1,015	1,019	2,034	No
Application 6	464	457	921	No
Application 7	685	695	1,380	No
Application 8	440	429	869	Yes
Application 9	5,597	5,691	11,288	Yes
Application 10	652	184	836	Yes
Application 11	861	832	1,693	Yes
Application 12	594	539	1,133	Yes
Application 13	2,603	3,046	5,649	Yes
Application 14	1,654	1,630	3,284	Yes

Appendix E: Instrument for Recording Study Variables

application	time1	time2	treatment
Application 1			0
Application 2			0
Application 3			0
Application 4			0
Application 5			0
Application 6			0
Application 7			0
Application 8			1
Application 9			1
Application 10			1
Application 11			1
Application 12			1
Application 13			1
Application 14			1