UNIVERSITY OF NAIROBI

SCHOOL OF COMPUTING AND INFORMATICS

A MULTI-DIMENSIONAL MODEL FOR DETERMINING SUSCEPTIBILITY TO UNINTENTIONAL INSIDER THREATS: THE CASE OF SOCIAL ENGINEERING THROUGH PHISHING

PAULA MWIKALI WASUA MUSUVA

A DOCTORAL THESIS
SUBMITTED IN PARTIAL FULFILMENT OF THE
REQUIREMENT FOR THE AWARD OF THE DEGREE OF
DOCTOR OF PHILOSOPHY IN INFORMATION SYSTEMS,
SCHOOL OF COMPUTING AND INFORMATICS,
UNIVERSITY OF NAIROBI

2019

COPYRIGHT

DECLARATION

I declare that this thesis is my original work and that it has not been presented for a degree in any other university.

Signature: …………………………………………… Date ……….……………
Paula Mwikali Wasua Musuva
P80/92723/2013

We declare that this thesis has been submitted for examination with our approval as supervisors:

Signature: …………………………………………… Date ……….……………
Katherine W. Getao, PhD, EBS
Chief Executive Officer,
ICT Authority, Kenya

Signature: …………………………………………… Date ……….……………
Christopher Kipchumba Chepken, PhD
Lecturer, School of Computing and Informatics,
University of Nairobi

ABSTRACT

A MULTI-DIMENSIONAL MODEL FOR DETERMINING SUSCEPTIBILITY TO UNINTENTIONAL INSIDER THREATS: THE CASE OF SOCIAL ENGINEERING THROUGH PHISHING

Paula Mwikali Wasua Musuva
Doctoral Thesis: 237 pages, 8 appendices, 76,882 words
School of Computing and Informatics
University of Nairobi

Many of the information security incidents that make headlines around the world are perpetrated by authorized users of the information systems. These users are commonly referred to as insiders. The Unintentional Insider Threat is posed by insiders who inadvertently compromise information systems. Literature shows that the Unintentional Insider Threat is under researched and should be the focus of current insider threat research. One predominant case of Unintentional Insider Threat is social engineering particularly through phishing. There is need for a unified multi-dimensional theoretical model that facilitates an understanding of the Unintentional Insider Threat phenomenon. This research is a response to this gap.

The presented multi-dimensional theoretical model is grounded on the Elaboration Likelihood Model and Protection Motivation Theory. In addition, it is developed after evaluating 62 research articles on the Unintentional Insider Threat. The model presents: 1 dependent variable, 22 independent variables and 12 control variables. This model is then validated using data from an empirical study that is guided by the realist, positivist and objective ontological and epistemological views; using a deductive research approach. Quantitative data is collected by staging a naturalistic experiment which presents a real-life social engineering phishing attack. This is after gaining approvals for the research from an institution's research board (IRB) and its administration. This allows study participants to be observed without alerting them on the ongoing research, therefore, providing data with high ecological and external validity. Participants are then requested to fill in a cross-sectional survey in order to measure latent constructs and variables that were not directly observed. Data is analyzed using Structural Equation Modeling (SEM) because the technique allows for all the variables and relationships to be tested in their entirety; and accommodates latent constructs in the model analysis. A total of 192 cases are analyzed from an effective sample size of 241 persons who participated in the experiment giving a 79.67% response rate. A total of 22 hypotheses are tested. Of these, 10 are supported while 12 are not supported by the provided model specification and sample dataset. The model is able to explain 41.4% of the Elaboration variance, 43.1% of Threat Detection variance, 19.1% of Threat Avoidance variance and more importantly 28.7% of Unintentional Insider Threat outcome variance and performs better than models presented in other studies.

This study makes several contributions to theory, knowledge, policy and practice. It presents a unified theoretical model that gives a multi-dimensional understanding of the Unintentional Insider Threat phenomenon from demographic, organizational, insider and attack factors. This model can be used to provide a theoretical grounding in the study of various unintentional insider threats and can also be comparatively applied by other researches in different contexts. The body of knowledge is extended in the testing and analysis of 22 hypotheses and discussion of the findings. The various factors presented in the multi-dimensional model encourage policy makers to address the Unintentional Insider Threat not only using technology but also through addressing psychological and sociological imperatives. Recommendations for policy and practice show that organizations should invest in measures that equip users with the ability to detect threats; particularly through their knowledge on detection cues and high determinants of trust. In addition, efforts must be taken to increase cognitive elaboration so as to intentionally counter factors that try to diminish insider’s ability to examine deceptive scenarios.

DEDICATION

To my loving family and to the academic posterity.

ACKNOWLEDGEMENTS

The Almighty God, Alpha and Omega of all things: thank you for seeing me through a difficult journey; for giving me the faith, resources and perseverance for this work.

My loving family: thank you for your prayers, support and encouragement.

The nannies: thank you for freeing me up to attend to this research.

Supervisors Dr. Katherine Getao, EBS and Dr. Christopher Chepken: thank you for making time, for your guidance and critique that improved my work.

University of Nairobi community, specifically PhD coordinator Prof. Peter Waiganjo Wagacha, Board of Post Graduate studies, examiners and wider administration: thank you for all your nurturing and support to earn this achievement.

Dr. Angela Mwende and Dr. Catherine Kavata: thank you for guiding me through this entire process as a sister and friend.

German Academic Exchange Service (DAAD): thank you for awarding me the scholarship that enabled me to pay my tuition and to conduct this research.

United States International University – Africa community and colleagues: thank you for your support in data collection and analysis.

To all researchers whose published work provided a stepping stone for this research; thank you for your mentorship and guidance.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

APT	Advanced Persistent Threat
APWG	Anti-Phishing Working Group
AVE	Average Variance Extracted
CERT	Computer Emergency Response Team
CFA	Confirmatory Factor Analysis
CFI	Comparative Fit Index
EFA	Exploratory Factor Analysis
ELM	Elaboration Likelihood Model
FTP	File Transfer Protocol
GFI	Goodness of Fit Index
GOF	Goodness-of-fit
HTTPS	Hyper Text Transfer Protocol Secure
ICT	Information and Communication Technology
IIT	Intentional Insider Threat
NFI	Normed Fit Index
PCFI	Parsimony Comparative Fit Index
PMT	Protection Motivation Theory
PNFI	Parsimony Normed Fit Index
RMSEA	Root Mean Square Error of Approximation
SEM	Structural Equation Modeling
TTAT	Technology Threat Avoidance Theory
UIT	Unintentional Insider Threat

CHAPTER 1: INTRODUCTION

1.1 Background of the Study

Information security incidents continue to make headlines around the world. Many of these incidents are facilitated by legitimate users of information systems - whether done deliberately or inadvertently. These legitimate users, who have authorized access to the systems, are referred to as insiders (CERT, 2013; Collins et al., 2016). Insiders present a unique challenge in the enforcement of information security in organizations. This is because insiders are in a position of trust and have the knowledge and capability to bypass information security controls (Verizon, 2015). In their study Cummings, Lewellen, McIntire, Moore, & Trzeciak (2012) found that 71% of the cases of insider attacks involved employees using authorized access.

An *insider***,** as defined by the Insider Threat Team at Carnegie Mellon University (CERT, 2013; Collins et al., 2016), is a current or former employee, contractor, business partner who has authorized access into the organizations systems. Bishop & Gates (2008) and Theoharidou, Kokolakis, Karyda, & Kiountouzis (2005) extend this definition to include any trusted entity that has the power to violate an information system's security policy. Insider action (or inaction) can compromise the security of the information system and therefore pose a threat known as the insider threat.

Information from recent survey reports show that insiders are the cause of a significant proportion of the current information security incidents organizations are experiencing (APWG, 2018; CA Technologies, 2018; proofpoint, 2019b, 2019a; Serianu & USIU-A, 2014; Verizon, 2018).

When conducting risk assessments many organizations focus on protecting their information assets from external intrusion but do not pay as much attention to the risks posed by trusted insiders. Yet studies show that the threat to the organization's information systems is greater from insiders than from outsiders (Carnegie Mellon University, 2013; Chinchani, Iyer, Ngo, & Upadhyaya, 2005; Flynn, Huth, Trzeciak, & Buttles, 2013).

Key findings from the 2015 U.S. State of Cybercrime Survey (PWC, CSO, CERT, & USSS, 2015) showed that 45% of those surveyed considered insider attacks to be more damaging than those by outsiders. An increase of 11% up from 34% in 2013. In 2013, the U.S. State of Cybercrime Survey (PWC, CERT, USSS, & CSO, 2013) had established that 53% of the organizations had experienced an insider incident. The data also revealed that the information security tools deployed in most organizations would be ineffective against insider threats since the insiders already had been granted access to the information system. In addition, these insiders understood the organization's systems and operations well enough to capture the most valuable information assets while bypassing detection.

According to Annual Cyber Security Reports in Africa (Serianu & USIU, 2017; Serianu & USIU-A, 2016) and Kenya (Serianu & USIU-A, 2014, 2015, 2017; Serianu, USIU-A, & Paladion, 2016), insider threats continue to be the biggest information security threat faced by most organizations. As an example, in 2013, 1.49 billion Kenya Shillings was stolen from bank accounts through schemes hatched by employees. In addition, a leading commercial bank reported an employee defrauding the bank of 60 million shillings. Another study released by Deloitte in East Africa covering an 18-month period from 2011 to June 2012, showed that East African financial institutions lost 4.06 billion Kenya Shillings to fraud. The Deloitte report highlighted that about 50% of the total fraud was committed with the help of the organization's employees (Mumo, 2012). In addition, the 2018 Global Economic Crime Survey Kenya Report (PwC, 2018) points out that twice as many respondents experienced crimes carried out by insiders as compared to those carried out by external actors.

Similarly, a study by Verizon (2018) with a database of over 53,000 security incidents from 65 countries showed that 28% of the attacks were by insiders. A similar study in 2013 attributed 69% of reported incidences to insiders with most of these incidents coming from accidental insider actions. A detailed analysis of confirmed insider incidents by Verizon shows that employees with little technical skills and lower cadre positions are commonly involved in the security incidents. These include cashiers, bank tellers and waiters who, for example, can skim payment cards or copy off account information to external parties and collaborate with them to execute fraud schemes. This is also confirmed by Cummings et al. (2012) where 80% of the insider attacks were carried by non-technical staff.

Pfleeger & Stolfo (2009) point out that it is hard to find credible data that describes the scope and impact of insider threats. Many organizations world-wide do not reveal details about the incidents they experience for fear of reputational damage. With this in mind, it is important to note that figures and statistics reported in various insider threat studies are understated because many cases go unreported.

1.1.1 Why study the Unintentional Insider Threat?

Two main categories of Insider Threats have been identified in literature (Andersen et al., 2004; CERT, 2013; Collins et al., 2016; Flynn et al., 2013; Greitzer et al., 2014; Silowash et al., 2012). The first category is of insiders who deliberately compromise information systems they have access to. In this category, the insider's actions (or inactions) are intentionally malicious, destructive, fraudulent and criminal in nature. Examples given by Cappelli, Moore, & Trzeciak (2012) and Collins et al. (2016) include the sabotage of information systems, theft of intellectual property and outright fraud. The second category is of insiders who accidentally, without malicious intent, compromise information systems. This category is known as the Unintentional Insider Threat (UIT).

Verizon in their 2015 Data Breach Investigations Report (Verizon, 2015) analyzed 79,790 incidents with 2,122 confirmed data breaches. They found that the top most frequently occurring incident classification pattern was incidents that result from insider error. The third most frequent occurring pattern was incidents resulting from insider misuse. AlgoSec (2013) reported that 40% of the security professionals surveyed considered accidental actions by users to be their greatest organizational risk.

A lot more research has gone into examining the malicious insider but little has gone into studying the Unintentional Insider Threat as highlighted by the Carnegie Mellon University CERT Insider Threat team in their foundational study of the UIT (CERT, 2013) and a follow-up study by Greitzer et al. (2014). A lot of the factors and theories relating to intentional malicious insider actions may not apply to unintentional insider threats (CERT, 2013; Greitzer et al., 2014; Luo, Zhang, Burd, & Seazzu, 2013).

This justifies the focus of this study on the unintentional insider threat since it is an area that is largely unexplored and under-researched.

1.1.2 Unintentional Insider Threat Taxonomy

This study focuses on the Unintentional Insider Threat; a topic that is under researched (CERT, 2013; Greitzer et al., 2014; Wang, Herath, Chen, Vishwanath, & Rao, 2012). This study's scope is further refined using the Insider Threat taxonomy provided by the Carnegie Mellon University CERT Insider Threat team (CERT, 2013). Homoliak, Toffalini, Guarnizo, & Elovici (2018) point out that the work by the CERT Insider Threat Team is the most relevant resource when establishing the scope of insider threats and is derived from the analysis of over 1000 real case studies. This taxonomy, illustrated in Figure 1, shows how insider threats play out in the wider organizational information security threat domain context. The term "task failure" as defined by the CERT Insider Threat team refers to the outcome of incorrect information processing and their taxonomy is based on the Trust Theory (Castelfranchi & Falcone, 2010; Urbano, Rocha, & Oliveira, 2013) .The taxonomy outlines seven negative impacts, colour-coded red in the diagram, that result from task failure, namely: malicious outsider attack, existential failure, engineering failure, malicious insider attack, outsider collusion attack, social engineering attack and human failure.

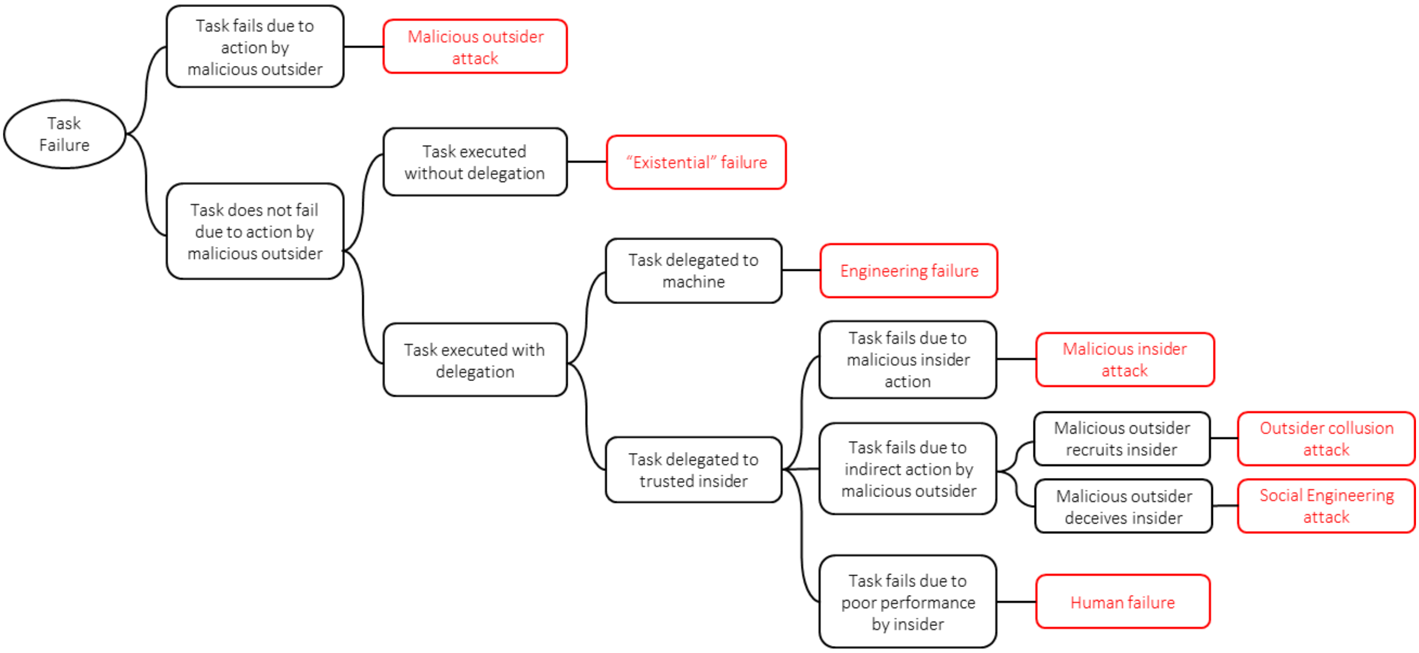

Task Failure
Task fails due to action by malicious outsider
Malicious outsider attack
Task does not fail due to action by malicious outsider
Task executed without delegation
"Existential" failure
Task executed with delegation
Task delegated to machine
Engineering failure
Task delegated to trusted insider
Task fails due to malicious insider action
Malicious insider attack
Task fails due to indirect action by malicious outsider
Malicious outsider recruits insider
Outsider collusion attack
Malicious outsider deceives insider
Social Engineering attack
Task fails due to poor performance by insider
Human failure

Figure 1: Insider Threat Taxonomy (CERT, 2013)

In the taxonomy, the first points of failure consider whether a malicious outsider action is involved. If not, the responsibility for successful information processing rests either with an automated machine or with a trusted insider. There are two specific scenarios where information processing delegated to a trusted insider fails due to the Unintentional Insider Threat. These are: (1) tasks that fail due to indirect action by a malicious outsider who deceives an insider and (2) those that fail due to poor performance by insider. The first category relates to social engineering while the second category relates to what is termed as human failure and these two categories are explored in more detail hereafter.

Unintentional Insider Threats from Human Failure

CERT (2013) Insider Threat Taxonomy breaks down the category of Unintentional Insider Threats (UITs) that result from human failure into four sub-groups. The first UIT sub-group is accidental disclosure of confidential information; for example, a system administrator who posts router configurations on a discussion forum when troubleshooting a network failure. Another example is an email with sensitive information being sent to wrong recipients. Verizon (2015) also highlight an example where organizations host poorly secured file servers containing repositories of confidential information (such as login information, medical records, legal agreements, project documents) that can be accessed on the internet with little effort.

The second UIT sub-group is through the accidental introduction of malware, such as viruses and spyware, into organizational systems. This could happen, for example, when an unsuspecting employee picks a flash drive on a parking lot and inserts it into their computer to check its contents. The flash drive in this case would have been infected with malware and planted somewhere an employee would be likely to find. Verizon (2015) points out that 35% of end users are vulnerable to such attacks.

The third UIT sub-group captures employees who do not dispose confidential information properly, for example, throwing away company procedure manuals without shredding them or computing devices, storage media such as old tapes or hard disks without securely wiping them.

The fourth UIT sub-group includes employees who lose portable computing equipment such as laptops, mobile phones or even hard disks that have confidential organizational data. The loss of these portable devices can be particularly devastating if they belong to high-ranking personnel who carry highly confidential organizational information.

Unintentional Insider Threats from Social Engineering

The other category of unintentional insider threats is termed social engineering and is the focus of this research. Social engineering is the use of manipulation tactics by malicious outsiders to get unsuspecting insiders to compromise an organization's information security, such as by providing access to confidential information or access to protected information systems (Luo, Brody, Seazzu, & Burd, 2011). This confluence of malicious outsider and non-malicious insider is the distinguishing feature pointed out by the CERT (2013) Insider Threat Taxonomy.

Hackers have demonstrated the use of simple social engineering techniques to bypass technical information security controls with relative ease. Kevin Mitnick, one of the most prolific hackers of the 20th Century, chronicles his tales of hacking through social engineering in his Book "The Art of Deception" (Mitnick & Simon, 2002). He explains that companies who invest in the best information security systems are still completely exposed due to social engineering. In fact, Algarni (2019) point out that hackers succeed even in organizations that state that their employees have been made aware of social engineering tactics.

This points out that the unintentional insider threat is an information security challenge that cannot be addressed by technology alone (Kandias, Mylonas, Virvilis, Theoharidou, & Gritzalis, 2010; Luo et al., 2011; Martinez-Moyano, Conrad, & Andersen, 2011; Schneier, 2000). Regrettably, organizations have placed a premium on addressing information security threats using technology without giving as much attention to other controls (Luo et al., 2011; Ophoff, Jensen, Sanderson-Smith, Porter, & Johnston, 2014).

Tetri & Vuorinen (2013) in their review of the literature were able to identify 24 attack vectors that can be employed to conduct social engineering attacks. Examples of attack vectors identified include dumpster diving, impersonation, pretexting, manipulation, phishing, tailgating and shoulder surfing. In addition, Tetri & Vuorinen (2013) were able to synthesize 3 distinct characteristics of social engineering attacks regardless of which of the 24 or more attack vectors was used. The 3 characteristics that make up a social engineering attack are: persuasion, fabrication and data gathering.

The first, persuasion, aims at getting insiders to comply with inappropriate requests made by a malicious outsider. Persuasion is commonly achieved through manipulating the emotions of the insider with relation to fear, greed or trust. The second, fabrication, aims at providing deceptive cues to the insider to dupe them as to what is actually taking place. Fabrication aims at giving legitimacy to the request made by the malicious outsider by providing symbols of legitimacy that the insider expects, for example, falsified identification badges and logos. The third characteristic, data gathering, is the crown jewel of social engineering attacks. It aims at getting the insider to perform an action that eventually compromises the security of an information systems. This could be by getting the insider to provide sensitive and confidential information that can give the malicious outsider access into the information system. It could also be by getting the insider to install malware that gives the malicious outsider control of the information system.

1.1.3 Phishing as an Unintentional Insider Threat

One unintentional insider threat that is highly prevalent and that manifests each of the characteristics of social engineering attacks is phishing. Phishing is described by the Anti-Phishing Working Group (APWG, 2018) as a criminal attack that uses deception over a technical medium in order to get users to give out their identity data, login credentials and other confidential information. The deception aims at getting the user to think that the communication is a legitimate request for their confidential data. Another way to describe phishing is simply ‘fishing’ for data (James, 2005). This is the use of social deception (the fishing bait) with the aid of communication technologies such as apps, email or websites (the fishing rod) in order to compromise the security of an information system (the catch).

The most common vector for delivering phishing attacks is email (James, 2005; Kumaraguru, Rhee, Acquisti, et al., 2007; Verizon, 2018) because it provides a way to reach large numbers of people with little effort. In addition, once an email is delivered to a user’s inbox it has crossed the boundaries of the external perimeter defenses and is now inside an organization’s network; thus, making it a very effective way of compromising information systems from inside the organization. Phishing emails are also used to deliver malware onto a user’s system that can then harvest confidential information and automate the attack process from within a local network.

Research by Verizon (2015, 2016, 2017), Fire Eye (2015, 2017) and Mandiant (2004, 2010), on recent cases of Advanced Persistent Threat involving Crimeware and Cyber-Espionage, show that a common technique of compromising organizations is by delivering phishing emails to targeted individuals. This phishing technique of crafting attacks to fit targeted individuals is called spear phishing. The spear phishing email is often well crafted to be relevant to the recipient and also appears to come from a legitimate sender, such as a colleague or company executive, because of a spoofed e-mail address.

Cases of phishing attacks are still on the rise despite a long history of phishing campaigns dating back to 1995 (James, 2005). Verizon (2018) established that 93% of the data breaches they examined involved phishing and manipulation of unsuspecting users using false messages. The Anti-Phishing Working Group report (APWG, 2017) on the fourth quarter of 2016 reported an increase of 65% in the number of phishing attacks compared to those reported in 2015. In addition, a trend analysis of phishing attacks since 2004 shows a 5,753% increase over a 12-year period. The previous report for the first quarter of 2016 (APWG, 2016) showed a 250% increase in the number of unique phishing websites since the last quarter of 2015. An increase by 250% in a period of six months is huge. PhishTank, another organization that monitors cases of phishing, reports that there were 4.5 million reported phishing sites in October 2016 and 42,788 of these were confirmed as active phishing sites (PhishTank, 2016).

Research by Cyveillance (2015) on the cost of phishing shows that phishing attacks are estimated to result in losses of 5.9 billion US dollars annually. News headlines in August 2016 (Barth, 2016; BBC News, 2016) highlighted a criminal network led by a 40 year old Nigerian man called “Mike” that scammed individuals and companies off 60 million US dollars through email scams and phishing malware. Previous research done by Hernandez, Regalado, & Villeneuve (2015) on Nigerian scammers show consistent use of email-based social engineering to defraud businesses of millions of dollars.

Investigative reports on allegations of Russia’s involvement in the 2016 elections in the United States of America (Fire Eye, 2017) also show compromise through spear-phishing emails targeted at key staff in the Democratic Party.

In April 2016, the hacktivist group Anonymous posted 1 Tera Byte (TB) of sensitive data from Kenya’s Ministry of Foreign affairs on the dark web. After the disclosure of the breach, Kenya’s ICT Cabinet Secretary explained that the hackers succeeded in gaining access to the ministry’s data through phishing. An email circulated by the head of IT dated 4th August 2015 (several months before the attack) tried to alert staff on the phishing attempts being sent by people impersonating the ICT administrator (Cimpanu, 2016; Obulutsa, 2016; Waqas, 2016).

In December 2018 going into January 2019, the Communications Authority of Kenya (CA) posted a cybersecurity advisory through the local media regarding the Emotet malware that targeted online banking and e-payment systems. The malware was spreading through phishing links and email attachments that appeared to be bank account alerts, payment notifications and invoices that had legitimate branding from affected institutions (Nyayieka, 2019a, 2019b; Odhiambo, 2019; Osongo, 2019). The National Computer Incident Response Team Coordination Centre (National KE-CIRT/CC) had established that 11 local institutions had been affected by the malware by the close of December 2018. A few weeks previously, Kenya’s National Cyber Centre (NCC) had released the cyber-attack statistics for July to September 2018 and had shown that cyber-attacks had increased by 11.76% to a total of 3.8 million incidents from 3.4 million in the previous quarter (Nyayieka, 2019b; Obura, 2018). These cases demonstrate that phishing is still a real active threat to users and a growing concern for organizations today and it must be addressed.

1.1.4 Previous Studies on Unintentional Insider Threats

Various empirical studies have tried to address the Unintentional Insider Threat. Many have examined the use of technology to prevent or detect attacks (Bose & Leung, 2007; Dhamija & Tygar, 2005; Fette, Sadeh, & Tomasic, 2007; Jakobsson & Myers, 2006; Miller & Wu, 2005; Wu, Miller, & Garfinkel, 2006; Zhang, Egelman, Cranor, & Hong, 2007). However, a review of these technological measures has revealed that they are inadequate in protecting users (Chuenchujit, 2016). A number of reasons have been proposed to explain why technology measures fall short in protecting users from unintentional threats. First, many tools have been found not to operate correctly or not to have good detection accuracy (Egelman, Cranor, & Hong, 2008; Wu et al., 2006; Zhang et al., 2007). Second, attackers are constantly looking for ways to make these technological measures ineffective and this ends up being an arms race (Downs, Holbrook, & Cranor, 2007). Third, even if the measures are effective, they rely on human beings to implement and use them correctly (Aytes & Connolly, 2004; Aytes & Conolly, 2003). Some users ignore warnings that signal information security compromise (Dhamija, Tygar, & Hearst, 2006; Wu et al., 2006).

The other group of studies have tried to address unintentional insider threats by focusing on the human factor. A careful examination of these studies has revealed seven categories of human factors. These are: *(1) lack of knowledge* (Dhamija et al., 2006; Downs, Holbrook, & Cranor, 2006; Downs et al., 2007; Jakobsson, Tsow, Shah, Blevis, & Lim, 2007; Vishwanath, Herath, Chen, Wang, & Rao, 2011) *(2) lack of effective training and awareness* (Kumaraguru et al., 2009; Kumaraguru, Sheng, Acquisti, Cranor, & Hong, 2008; Kumaraguru, Rhee, Acquisti, et al., 2007; Kumaraguru, Rhee, Sheng, et al., 2007; Kumaraguru, Sheng, Acquisti, Cranor, & Hong, 2007; Sheng et al., 2007) *(3) effective persuasion and deception techniques used by attackers* (Jagatic, Johnson, Jakobsson, & Menczer, 2007; Jakobsson et al., 2007; Luo et al., 2013; Rusch, 1999; Workman, 2007, 2008a, 2008b) *(4) poor perception of negative consequences* (Aytes & Connolly, 2004; Downs et al., 2007; Workman, 2007) *(5) personality based factors* such as personality traits, risk propensity, need for cognition (Kumaraguru, Rhee, Sheng, et al., 2007; Luo et al., 2013; Sheng, Holbrook, Kumaraguru, Cranor, & Downs, 2010) *(6) cognitive processing of the threat* (Luo et al., 2013; Vishwanath et al., 2011; Workman, 2008b) *(7) demographic factors* such as: age, gender, level of education, area of specialization, years on the internet, emails received per day, hours spent online per day, online activities engaged in such as online shopping and banking, computer skill, prior victimization and prior training (Dhamija et al., 2006; Downs et al., 2006, 2007; Jagatic et al., 2007; Kumaraguru et al., 2009; Kumaraguru, Rhee, Acquisti, et al., 2007; Kumaraguru, Rhee, Sheng, et al., 2007; Kumaraguru, Sheng, et al., 2007; Sheng et al., 2010, 2007, p.; Workman, 2007, 2008b).

These factors have been tested empirically through: *field studies* (Downs et al., 2007; Luo et al., 2013; Vishwanath et al., 2011; Workman, 2007, 2008b; Workman, Bommer, & Straub, 2008), *field experiments* (Jagatic et al., 2007; Kumaraguru et al., 2009, 2008; Sheng et al., 2010; Workman, 2008a), *lab studies* (Dhamija et al., 2006; Downs et al., 2006; Jakobsson et al., 2007) and *lab experiments* (Kumaraguru, Rhee, Acquisti, et al., 2007; Kumaraguru, Rhee, Sheng, et al., 2007; Kumaraguru, Sheng, et al., 2007; Sheng et al., 2007).

Findings show that the more knowledgeable users are on the threat and on detection techniques, the less susceptible they are to unintentional threats (Downs et al., 2007; Kumaraguru, Rhee, Sheng, et al., 2007; Kumaraguru et al., 2008; Sheng et al., 2010, 2007). The term susceptibility refers to the likelihood of falling victim to the threat (Algarni, 2019). Many users succumbed to unintentional threats because they were not aware of the threat and also did not know how to correctly interpret trust indicators or deception cues presented to them. When users are taken through some training and awareness they become less susceptible to unintentional threats (Kumaraguru et al., 2009; Kumaraguru, Rhee, Acquisti, et al., 2007; Kumaraguru, Rhee, Sheng, et al., 2007; Sheng et al., 2010) ..

Research has also shown that the more persuasive and deceptive an attacker is, the more users are likely to succumb to unintentional threats. Jakobsson et al. (2007) noted that a high degree of attack personalization increased trust. Jagatic et al. (2007) demonstrated that by sending spear phishing emails that were customized from social network profiles, attack success rates were as high as 72% as opposed to generalized phishing emails at 16% success. Dhamija et al. (2006) found that well designed phishing sites fooled 90% of the participants. Rusch (1999) presents Caldini's six principles of influence and persuasion (authority, scarcity, liking and similarity, reciprocation, commitment and social proof) in the context of social engineering. Workman (2007, 2008b) was able to study 4 of these principles (authority studied as obedience, liking and similarity studied as trust, commitment and reactance) and shows that when they are used people succumb more to the threat.

Prior research also shows that those who have a lower perception of threat severity and vulnerability are more likely to succumb to unintentional threats (Workman, 2007; Workman et al., 2008). Another approach of rating negative consequences of online threats used by Downs et al. (2006, 2007) showed that if users perceived higher severity they were also less likely to interact with legitimate sites.

People's personality also predisposes them in certain ways to succumb to unintentional threats. Kumaraguru, Rhee, Sheng, et al. (2007) and Sheng et al. (2010) found that people who were more impulsive and risk taking were likely to succumb to phishing. Luo et al. (2013) also notes that people who have a high need for cognition are more likely to engage mentally in objectively processing a threat and less likely to rely on deceptive cues in determining legitimacy. They are therefore less likely to succumb to unintentional threats.

Other studies used the Elaboration Likelihood Model (Vishwanath et al., 2011) and the Heuristic Systematic Model (Luo et al., 2013) to examine cognitive factors that affect people's susceptibility to the threat. They found that users who were more cognitively involved in evaluating the threat characteristics were less likely to succumb to the threat. The users who responded out of urgency or habit were more likely to succumb to the threat.

With regard to demographic factors, some studies have been unable to demonstrate a relationship to unintentional threats (Dhamija et al., 2006; Kumaraguru, Sheng, et al., 2007). However, other studies showed that women were more susceptible than men and that younger people within the 18-35 year age group are more susceptible (Jagatic et al., 2007; Kumaraguru et al., 2009; Sheng et al., 2010). In addition, those who received many emails were more likely to respond to phishing emails out of habit (Luo et al., 2013). Users who were tech-savvy and had a background in engineering or computing were less likely to succumb to phishing (Jagatic et al., 2007; Kumaraguru et al., 2009). Sheng et al. (2010) established that the reason why women succumbed more was because they had less technical knowledge and training than men. Similarly, younger people were more susceptible because they had less education, fewer years on the internet, little exposure to anti-phishing training and poorer risk perception than older age groups. Those who had been prior victims had a higher severity perception. Downs et al. (2006) and Downs et al. (2007) found that those who had prior encounters with spoofed illegitimate sites were less likely to succumb to phishing.

1.2 Research Problem

Empirical evidence and theoretically-grounded models are scarce and not well developed in the unintentional insider threat literature (Greitzer et al., 2014; Jones & Towse, 2018; Luo et al., 2013; Tetri & Vuorinen, 2013; Vishwanath et al., 2011; Wang et al., 2012). Greitzer et al. (2014) state that the unintentional insider threat research topic is largely unrecognized and calls for more research in this area. (Luo et al., 2013; Wang et al., 2012) state that there is great need for unintentional insider threat research that investigates the theoretical underpinning of phishing susceptibility. Tetri & Vuorinen (2013) and Vishwanath, Herath, Chen, Wang, & Rao (2011) state that missing from existing literature is a single comprehensive model that examines the unintentional insider threat from multiple perspectives. Furthermore, Jones & Towse (2018) state that existing behavioural models are still not well developed.

Earlier models focused on examining insider-based demographic and human factors that made unintentional insider threats succeed (Kumaraguru et al., 2009, 2008; Sheng et al., 2010). Later models included the analysis of various attack factors and their influence on unintentional insider threat susceptibility (Algarni, 2019; Vishwanath, Harrison, & Ng, 2018; Williams & Polage, 2019).

Greitzer et al. (2014) and CERT (2013) show the need for a multi-dimensional approach when examining antecedent factors to the unintentional insider threat phenomena. They propose the following dimensions of factors: *(1) demographic factors* that characterize and describe individuals; *(2) organizational factors* that consider approaches taken to effectively protect users from the threat either using technology, policy, process or training of users; *(3) human factors* that take into account insiders' knowledge and awareness of the threat, risk tolerance, personality traits, cognitive processing of an attack in progress and the ability to properly use countermeasures to protect themselves. In addition, Tetri & Vuorinen (2013) propose a another factor dimension *(4) attack factors* that consider characteristics of the attack based on persuasive and deceptive techniques.

Tetri & Vuorinen (2013) and Vishwanath et al. (2011) point out that there is no theoretical model that presents such a multi-dimensional approach. In addition, there is no study that empirically examines the effects among all these four multi-dimensional factors. Having these factors analyzed together in an empirical study that proposes a unified theoretical model would make it possible to examine combined cause-and-effect relationships and interactions among factors and therefore provide a clearer and more comprehensive understanding of the unintentional insider threat phenomenon. This research is a response to this knowledge gap. It develops a unified multi-dimensional theoretical model that explains why insiders succumb to unintentional insider threats with an examination of the interactions between these four multi-dimensional factors. It also validates this model using social engineering as a particular case of the unintentional insider threat.

1.3 Research Objectives

This research seeks to develop and validate a unified multi-dimensional theoretical model for determining susceptibility to the unintentional insider threat to information systems security.

1.3.1 Specific Objectives

This research aims to meet the following specific objectives:

Objective 1: To establish a theoretical foundation for the factors that contribute to the unintentional insider threat to information systems security.

Objective 2: To develop a unified multi-dimensional theoretical model that explains susceptibility to the unintentional insider threat to information systems security.

Objective 3: To validate the unified multi-dimensional theoretical model using empirical data and appropriate statistical methods.

1.4 Scope

This research studies the unintentional insider threat from the most prevalent and active case of social engineering known as phishing delivered through emails (APWG, 2016; James, 2005; Kumaraguru, Rhee, Acquisti, et al., 2007; Mandiant, 2004, 2010; Verizon, 2015, 2018). The foundational unintentional insider threat taxonomy by the Carnegie Mellon University CERT Insider Threat Team (CERT, 2013) classifies social engineering as an attack that involves indirect action by a malicious outsider who deceives an insider. Research by Homoliak et al. (2018)

recognize this taxonomy that is maintained by the CERT division as the most appropriate way of establishing the scope of insider threats. This taxonomy distinguishes social engineering from other unintentional insider threats that result from human failure and poor performance. Therefore, social engineering stands out uniquely as an attack category and not an accidental occurrence. This qualifies social engineering as an appropriate phenomenon for study.

In their seminal work, Tetri & Vuorinen (2013) were able to show that social engineering attacks can be characterized using 3 distinct features regardless of the specific vector used to perpetrate the attack. They came to this conclusion after analyzing 24 different attack vectors employed in social engineering; some of which were: phishing, dumpster diving, eavesdropping, impersonation, pretexting, manipulation, tailgating and shoulder surfing. The three features that characterize social engineering attacks are: (1) persuasion (get insiders to fulfill inappropriate requests that may compromise information security); (2) fabrication (provide deceptive cues to the insider in order to prove legitimacy); and (3) data gathering (grant access to the information system to the malicious insider enabling the capture of sensitive and confidential information.

Phishing provides an ideal case for investigating social engineering as an unintentional insider threat because it manifests all these three dimensions. In addition, studying it is very relevant because it is a highly prevalent attack as highlighted in recent cybersecurity reports.

The Verizon (2017) Data Breach Investigations Report showed that 43% of all breaches investigated were due to social engineering. In addition, 92% of these social engineering attacks were perpetrated through phishing. This figure increased to 98% in the 2018 report (Verizon, 2018) with email still being the most common (96%) technique of delivering the attack.

Cases reported in recent cyber security intelligence reports (Fire Eye, 2015, 2017; Mandiant, 2004, 2010) highlight sophisticated and devastating attacks termed as the Advanced Persistent Threat (APT). These attacks are often carried out by organized groups with expert skills and a wealth of resources (sometimes funded by nation-states such as China and Russia). Most of their victims are compromised through targeted spear-phishing emails that install malware on their systems and siphon high-value confidential information.

In Kenya, a widely publicized information security breach by Anonymous in April 2016 also was perpetrated through phishing. Kenya’s ICT Cabinet Secretary explained that the attackers gained access to the Ministry of Foreign Affairs servers and data through successful phishing attacks targeted at the ministry’s staff. Over 1 Tera Byte of sensitive data was leaked on the dark web causing great embarrassment and reputational damage to the government.

Picking a specific unintentional insider threat (social engineering) and subsequently a specific attack vector (phishing) allows this study to focus to a greater level of detail. In order to ensure that the results of the research are still generalizable to the larger case of unintentional insider threats, care has been taken to ensure all the characteristics of the unintentional insider threat are addressed in the study. These characteristics as outlined by Tetri & Vuorinen (2013) are: persuasion, fabrication and data gathering.

1.5 Significance of the Study

This study provides a holistic and comprehensive understanding of the unintentional insider threat by examining demographic, organizational, human and attack factors in a multi-dimensional theoretical model that is validated through an empirical study. Previous studies on unintentional insider threats are mostly of an empirical nature and lack a grounding in theory (Luo et al., 2011; Tetri & Vuorinen, 2013; Vishwanath et al., 2011; Workman, 2007). In addition, these studies have focused on specific dimensions of the phenomenon and have not presented a holistic multi-dimensional understanding of the unintentional insider threat. The outcomes of this study provide significant contributions to academia, information security practice and policy.

The academia and research community benefits from the articulation of a multi-dimensional model with robust theoretical foundations and provision of empirical findings in an area that has been described as largely under-researched (CERT, 2013; Greitzer et al., 2014) and poorly grounded in theory (Luo et al., 2011; Tetri & Vuorinen, 2013; Vishwanath et al., 2011; Wang et al., 2012; Workman, 2007). This study addresses knowledge gaps in critical areas of theory and extends empirical findings in the existing body of knowledge by examining how the various multi-dimensional factors relate in a unified model.

This study also challenges information security practice to address the unintentional insider threat from a more comprehensive perspective. Many organizations have been found to focus their risk assessments on external intrusion and have deployed controls that are ineffective in addressing unintentional insider threats (Carnegie Mellon University, 2013; Chinchani et al., 2005; Flynn et al., 2013). In addition, many in the information security practice have focused on the use of technology without giving much attention to other factors (Luo et al., 2011). Focus on technology and ignoring other factors has proved to be the Achilles heel in otherwise highly secured information systems (Mitnick & Simon, 2002). It is important that organizations examine other solutions that take into account demographics, human factors and attack characteristics as this research will explore. The U.S. State of Cybercrime report by PricewaterhouseCoopers (PWC et al., 2013) showed that 33% of organizations have no formal plan or approach to mitigate insider threats or even investigate cases involving an insider. The findings in this study can help organizations formulate effective approaches to mitigate unintentional insider threats from multi-dimensional perspectives.

The particular case of social engineering through phishing focuses on a very prevalent and devastating case of unintentional insider threats that is directly costing organizations billions of dollars in losses. Cyveillance (2015) pegged the cost of losses at 5.9 billion US dollars annually. Reports in August 2016 (Barth, 2016; BBC News, 2016) on business email scams perpetrated by a ring of fraudsters led by a 40 year old Nigerian showed that companies lost over 60 million US dollars. This study provides an empirical study that highlights key vulnerabilities that predispose organizations to attack and subsequent financial losses. The findings of this research can assist organizations comprehensively address the unintentional insider threat and plug-in loopholes that are costing them vital resources and business advantage.

The results of this study will also help policy makers and regulatory bodies to develop the right policies in relation to information security controls that should be implemented to safeguard against insider threats. Cyber Security teams at national and international level have acknowledged that the unintentional insider threat is a serious challenge in securing cyberspace (Carnegie Mellon University, 2013; CERT, 2013). Devastating attacks have spread across countries and continents due to unintentional insider threat actions. One such case is the rise in ransomware attacks that has had devastating effects on a global scale in a matter of hours. The Verizon (2017) Data Breach Investigations Report shows that the trend in ransomware attacks is to deliver ransomware through phishing emails. Kenya's Cybersecurity Strategy and the National Cybersecurity Master Plan (Government of Kenya, 2012; Ministry of Information Communications and Technology, 2014) acknowledge that there is an increasing attack sophistication in the Cybersecurity landscape. Some of the attack vectors highlighted relate to unintentional insider threats through phishing and other social engineering techniques. Such efforts to outline national cybersecurity strategies, master plans, frameworks and policies would greatly benefit from the insights obtained from this study. Results of this study would ensure such efforts address critical factors that are often overlooked when securing information systems.

1.6 Definition of Terms

The following is a list of some commonly used terms in this thesis and their associated meanings:

- **Insider:** a current or former employee, contractor, business partner or other similar user who has authorized access into the organizations systems.
- **Intentional Insider Threat**: insiders who pose a danger to the security of information systems because of actions (or inactions) that are intentionally malicious, destructive, fraudulent and criminal in nature. These actions include the sabotage of information systems, theft of intellectual property and outright fraud.
- **Unintentional Insider Threat**: insiders who pose a danger to the security of information systems because of accidental actions, without malicious intent, compromise information systems. These actions include: accidental disclosure of confidential information, poorly secured servers by a system administrator, introduction of malware and clicking of harmful links on emails.

- **Phishing:** an attack where a user is tricked to submitting sensitive information, or installing malicious software by an attacker posing as a legitimate entity in an electronic communication such as email, social media post or chat.
- **Social Engineering:** the use of deception and manipulative tactics by an attacker in order to get insiders to grant them access to an organization and its information systems.
- **Susceptibility:** the likelihood of being responsive and falling victim to a targeted threat such as a social engineering attack through phishing

1.7 Outline of the Thesis

This thesis is written in six chapters.

Chapter one has set a background for this study by defining the key concepts relating to unintentional insider threats to information systems security and identifying knowledge gaps that need to be addressed. It has also outlined the research problem and stated the purpose of the research and its specific objectives. It has provided a scope and justification for the research by describing the significance of the study to academia, information security practice and policy makers.

Chapter two establishes the theoretical underpinning and empirical foundations for the study by examining the existing body of knowledge in relation to unintentional insider threats. The literature review is guided by the research objectives outlined in chapter one. Various factors are proposed based on what the existing literature says about the unintentional insider threat. Chapter two also presents a unified theoretical model that depicts antecedents and causal relationships to the unintentional insider threat. Constructs relating to this model are explained and operationalized with measures that can be examined through an empirical study.

Chapter three outlines the research design for the empirical study used to validate the multi-dimensional theoretical model. It explains the chosen research design philosophy by explaining the various ontological and epistemological considerations. It describes the research setting and the methodology followed to collect and analyze data. It discusses how sampling of the population was done, the development of the data collection instrument, data collection procedures and data analysis procedures.

Chapter four describes the various data analysis procedures undertaken and presents the findings of the research. It details results from various procedures, particularly in the following: data entry, coding, descriptive analysis, exploratory cluster analysis, exploratory factor analysis, confirmatory factor analysis, structural equation modeling and hypothesis testing.

Chapter five discusses the research findings in relation to the existing body of knowledge and highlights new knowledge that is generated. It examines the results in line with the research objectives set out at the beginning of the study. It also discusses the findings with the aim of affirming existing knowledge or highlighting new knowledge gleaned from the research.

Chapter six summarizes the research by outlining key findings, conclusions, contributions, implications, limitations and recommendations for future research.

The Appendix section provides the different detailed documents used during the study. These include the data collection letters, institutional review board approval and the data collection instruments that were used in the field and also detailed tables extracted during data analysis steps. In addition, copies of two journal publications extracted from this research are attached in the appendix.

CHAPTER 2: LITERATURE REVIEW

2.1 Introduction

This chapter will give a brief overview of relevant theoretical foundations and empirical studies that relate to this research on unintentional insider threats. The literature review is outlined in line with the research objectives set out for the study. First, the factors that contribute to the unintentional insider threat will be outlined from a theoretical and also empirical perspective in Section 2.2 and Section 2.3 respectively. Second, a conceptual theoretical model that depicts the relationship among the unintentional insider threat factors will be presented. The constructs used, their causal relationships and operationalized measures will be discussed in view of previous studies.

2.2 Theoretical Foundation

Various studies on intentional insider threats have examined a number of theories to understand why insiders deliberately misuse their organization's information systems and how to prevent them from doing so. The following is a list of theories and the studies that have explored the intentional insider threat perspective: *Theory of Planned Behaviour (TPB)* (Bulgurcu, Cavusoglu, & Benbasat, 2010; Herath & Rao, 2009b; Lee & Kozar, 2005); *Theory of Reasoned Action (TRA)* (Pahnila, Siponen, & Mahmood, 2007) ; *Rational Choice Theory (RCT)* (Bulgurcu et al., 2010); *General Deterrence Theory* (GDT) (D'Arcy, Hovav, & Galletta, 2009; Herath & Rao, 2009b; Pahnila et al., 2007; Siponen & Vance, 2010; Straub, 1990); *Protection Motivation Theory (PMT)* (Bojmaeh, 2015; Herath & Rao, 2009b; Johnston & Warkentin, 2010; LaRose, Rifon, & Enbody, 2008; Lee & Larsen, 2009; Pahnila et al., 2007; Tsai et al., 2016; Vance, Siponen, & Pahnila, 2012; Waleed, 2016; Woon, Tan, & Low, 2005; Workman et al., 2008); *Social Cognitive Theory* (LaRose et al., 2008; Rhee, Kim, & Ryu, 2009; Workman et al., 2008); *Social Comparison Theory* (Rhee, Ryu, & Kim, 2005); *Neutralization Theory* (Siponen & Vance, 2010); *Agency Theory* (Herath & Rao, 2009a); *Control Theory* (Boss, Kirsch, Angermeier, Shingler, & Boss, 2009); and *Theory of Cognitive Moral Development* (Myyry, Siponen, Pahnila, Vartiainen, & Anthony, 2009).

However, it has been found that the case of intentional insider threats is markedly different from the unintentional insider threats – particularly from a theoretical and empirical perspective (CERT, 2013; Greitzer et al., 2014; Luo et al., 2013). It has been argued that these theories cannot be directly applied to the study of unintentional insider threats (Liang & Xue, 2009, 2010).

First, these theories have been used to explain why the insiders intentionally fail to comply with organizational initiatives that prescribe security behaviours. They have examined situations where insiders have been instructed on the secure use of information systems, for example through information security policies; but have failed to do so. Workman et al. (2008) term this as the knowing-doing gap.

In contrast, the case of unintentional insider threats is different in that insiders may lack the awareness or knowledge of the expected security behavior. In many cases the insiders have not been trained or instructed on how to handle situations that pose a threat to the information system (Dhamija et al., 2006; Downs et al., 2006, 2007; Jakobsson et al., 2007; Vishwanath et al., 2011).

Second, these theories have taken the approach of examining acceptance behaviours as opposed to avoidance behaviours. Acceptance behaviours are associated with embracing or adopting virtuous information systems and practices. In contrast, the case of unintentional insider threats, particularly relating to social engineering, is associated with avoidance behaviours that dissuade users from malicious information systems and practices (Liang & Xue, 2009).

The Cybernetic Theory (Wiener, 1948) explains that human beings regulate their behaviours using what are known as feedback loops. Carver, 2006; Carver & Scheier (1982) used this concept of feedback loops to show a clear theoretical distinction between acceptance and avoidance behaviours. The acceptance behaviours present negative feedback loops that close the gap between a user's current state and a desired end state. However, avoidance behaviours present positive feedback loops that intend to widen the gap between a user's current state and an undesired end state.

With this distinction of acceptance and avoidance behaviours, it is important to examine theories that prescribe avoidance of malicious information systems and practices. Malicious attackers attempt to manipulate insiders to use malicious components or act in insecure ways (Grazioli, 2004). The intended behavioral outcome staged by a malicious attacker is not a secure behavioural outcome unlike the case of acceptance behaviours set by the organization. Additional consideration needs to be given before applying these theories that have been used to examine intentional insider threats to the case of unintentional insider threats.

This leads us to the third distinguishing factor which is deception. Deception is a key characteristic of unintentional insider threats. Deception occurs when there is misrepresentation by an opportunistic agent who intends to influence the behaviour of another target agent (Hyman, 1989; Johnson, Grazioli & Jamal, 1993; Russow, 1986). Those who fail to detect this deception make decisions and take actions based on the misrepresentation and thereby play into the attacker's snare. Key to understanding why users of an information system fall for unintentional insider threats is understanding why people fall for deception.

Johnson et al. (2001) explains that deception often succeeds because either; (1) the deceiver takes advantage of weaknesses in the way the target processes the information; or (2) the deceiver is aware of the target's detection efforts and acts in ways to frustrate them; or (3) the target lacks sufficient information to make a correct judgment and act in a secure manner. Therefore, secure behavioural response in the case of unintentional insider threats is largely determined by a person's ability to detect the deception, despite various deceptive and persuasive cues, and their ability to choose a response that will not compromise the security of the information system. Falling for an unintentional insider threat often stems from the inability to detect the threat, an inclination to believe an attacker's persuasive message and poor judgment resulting from a lack of knowledge or awareness.

With these distinguishing features in mind, it is necessary to examine unintentional insider threats using theories that relate to threat detection, threat avoidance, deception and persuasion. Liang & Xue, 2009 (2010) propose this approach by presenting the *Technology Threat Avoidance Theory (TTAT)*. They provide the threat appraisal and coping appraisal constructs and in addition propose other constructs namely; perceived threat, perceived avoidability, avoidance motivation and avoidance behaviour as the behavioural outcome. However, their work focuses on the threat detection and threat avoidance perspectives but does not examine the deception and persuasion perspectives. It can also be argued that the TTAT is not really a theory but rather a model that explains technology threat avoidance. Sutton & Staw (1995) explain that scientific theory is not about elucidating constructs, drawing diagrams or formulating hypothesis. It involves robust empirical evidence and arguments as to why certain phenomena have been observed and why the model is generalizable to a theory. Liang & Xue (2009) provide a good theoretical model but it is not supported by robust empirical evidence in field studies and neither does their work demonstrate that it is generalizable.

This research proposes theories relating to threat detection, threat avoidance, deception and persuasion that have been existed for a number of years and have also been examined in previous information systems studies such as the: *Interpersonal Deception Theory* (Vishwanath et al., 2011), *Theory of Deception* (Johnson et al., 2001; Vishwanath et al., 2011), *Elaboration Likelihood Model* (Rusch, 1999; Vishwanath et al., 2011; Workman, 2007, 2008b), *Heuristic Systematic Model* (Luo et al., 2013) and *Protection Motivation Theory* (Workman, 2007; Workman et al., 2008). (Cialdini, 2001) also explored deception techniques that attackers commonly use to manipulate insiders and they are called *Cialdini's Principles of Influence and Persuasion*. They include: authority, scarcity, liking and similarity, reciprocation, commitment and social proof. Additional factors proposed by (Bezuidenhout, Mouton, & Venter, 2010) and (Peltier, 2006) include strong affect/emotions, overloading, deceptive relationship and diffusion of responsibility.

2.2.1 Interpersonal Deception Theory

The Interpersonal Deception Theory (IDT) by Buller & Burgoon (1996) has been used in a previous study (Vishwanath et al., 2011) as a theory of interest in the case of unintentional insider threats. It analyzes deception when it takes place in interactive contexts, as illustrated in Figure 2, which are mostly face-to-face encounters. The sender (deceiver) and the receiver (deceived) are able to gauge each other’s responses and adapt their behaviour during the deception process. Therefore, the sender is able to strategically alter their message based on the responses they observe from the receiver (even if they are non-verbal) in order to carry out successful deception.

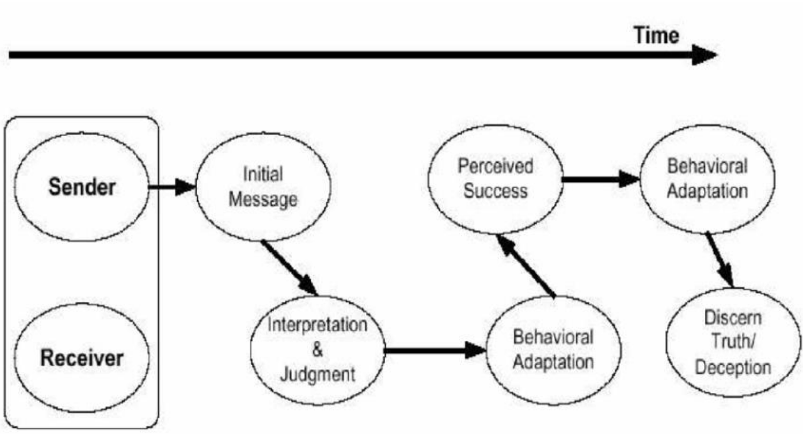


Figure 2: Interpersonal Deception Theory (Buller & Burgoon, 1996)

The Interpersonal Deception Theory may be useful when examining cases of unintentional insider threats delivered through active inter-personal engagement between an attacker and the insider. For example, when social engineering is handled through a phone conversation (a technique referred to as vishing). The key element here is the ability of the attacker to evaluate the responses from the insider in order to adapt their deception. However, in cases of unintentional insider threats where there is no interactive engagement, this theory may not be as suitable; as is our case on social engineering through phishing emails.

2.2.2 Theory of Deception

Another theory of interest is the Theory of Deception advanced by Johnson, Grazioli, Jamal, & Zualkernan (1992); Johnson, Grazioli, Jamal, & Berryman (2001); and Grazioli (2004). It has been commonly applied in various disciplines to understand how consumers of information detect deceptive communication.

It is largely similar and consistent with the Interpersonal Deception Theory which focuses on an individual's information processing during deception. However, it differs in 3 areas (Grazioli, 2004; Johnson et al., 2001; Vishwanath et al., 2011). First, the Interpersonal Deception Theory focuses mostly on the areas of communication and social psychology. However, the Theory of Deception has found use in wider disciplines and business contexts. Of particular interest are studies in the information and communication technology discipline that examine online deceptions occurring on the internet (Grazioli, 2004; Grazioli & Jarvenpaa, 2001; Vishwanath et al., 2011). Second, the Theory of Deception covers deceptions that have lower interactivity between the deceiver and target. It focuses on those that involve the evaluation of content as opposed to the high interactivity that the Interpersonal Deception Theory addresses. Third, the Theory of Deception does not focus on the interplay between the deceiver and the target. Rather, it focuses on the cognitive processing that occurs in the target when they are interacting with the deceptive communication. It examines the mental processing by the target and their ability to reason through the deception.

The Theory of Deception sets out four (4) processes in the detection of deception (Johnson et al., 2001, 1992) as illustrated in Figure 3. The first process is *activation*. Here the recipient pays attention to the deceptive message and evaluates it by picking deception cues that are inconsistent with the expectations of an authentic message. The second is *deception hypothesis generation*; where the recipient may try to come up with various hypothesis to explain the difference between what is expected and what is observed. Deception detection is only possible if the individual considers the possibility that the inconsistencies are due to deception. The third is *hypothesis evaluation*; where each of the generated hypothesis are analyzed against specific criteria and either accepted or rejected. Successful detection of deception at this stage requires knowledge on the domain and competencies at evaluating deceptive cues. The fourth and final stage is *global assessment*; where results of the hypothesis evaluation are amalgamated into one overall assessment of deceptiveness.

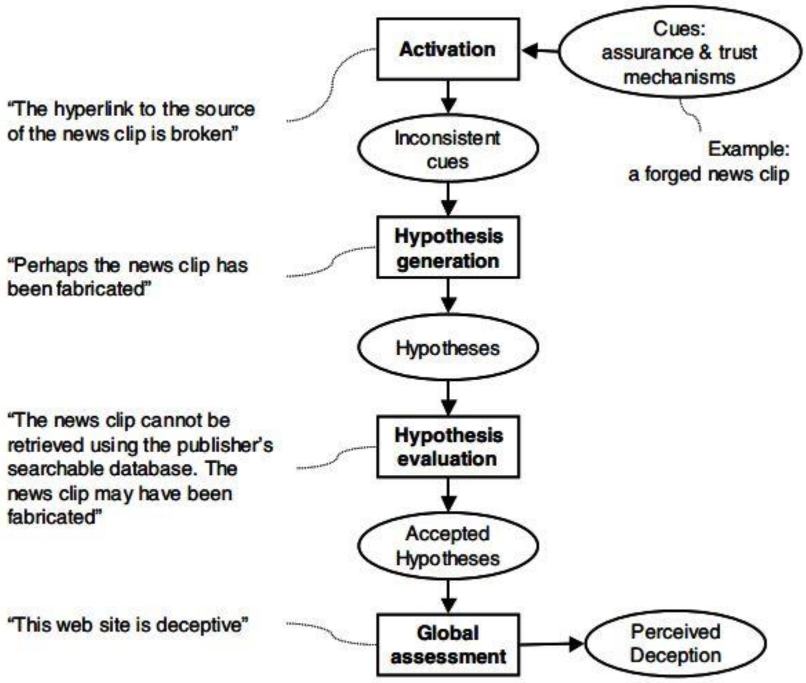


Figure 3: Theory of Deception (Grazioli, 2004; Johnson et al., 2001)

The Theory of Deception emphasizes the need for the recipient to have sufficient and competent knowledge regarding the domain in which the deception occurs; particularly of the deception techniques used by the deceiver and also the cues that can be used to detect the deception (Vishwanath et al., 2011).

This theory fits very well in the case of unintentional insider threats involving social engineering because it systematically guides the evaluation of the cognitive processes undertaken by insiders to identify gaps that lead to successful attacks. It also emphasizes the need to evaluate the insider's domain specific knowledge and their understanding of detection cues. Various studies have shown that these are key to understanding unintentional insider threats (Dhamija et al., 2006; Downs et al., 2006, 2007; Grazioli, 2004; Kumaraguru, Sheng, et al., 2007; Vishwanath et al., 2011).

One weakness in the Theory of Deception is its inability to distinguish different types of cues that could be evaluated when detecting deception. For example, some studies have shown that if insiders focus on persuasive cues they are more likely to fall for deception than if they focus on quality of the argument given or on threat detection cues (Luo et al., 2013; Vishwanath et al., 2011).

Another weakness is that the Theory of Deception does not address the influence that emotional factors have on the detection of deception. It only approaches deception from a rational thinking perspective (Grazioli, 2004).

In order to address these deficiencies, other theories are proposed to examine the case of unintentional insider threats; namely the Heuristic Systematic Model (Chaiken, 1980) and the Elaboration Likelihood Model (Petty & Cacioppo, 1986). Both the Heuristic Systematic Model (HSM) and Elaboration Likelihood Model (ELM) propose dual-processing modes for cognitive evaluation of persuasive communication. These dual-processing theories provide a fuller explanation compared to one-process approaches advanced by Theory of Deception and others such as Cognitive Dissonance Theory (Festinger, 1957) and Reactance Theory (Brehm, 1966) that have also been evaluated in unintentional insider threat research (Workman, 2007, 2008b).

2.2.3 Elaboration Likelihood Model Verses Heuristic Systematic Model

Both the Elaboration Likelihood Model (ELM) by Petty & Cacioppo (1986) and Heuristic Systematic Model (HSM) by Chaiken (1980) propose two cognitive processing modes during the evaluation of persuasive communication. The first cognitive processing mode is termed as "central" in ELM and "systematic" in HSM. It is characterized by a person's careful reasoned evaluation of the issue-relevant arguments presented by persuasive communication. The second cognitive processing mode is described as "peripheral" in ELM or "heuristic" in HSM and is characterized by low cognitive processing of the issue-relevant arguments. Instead, reliance is placed on simple peripheral cues to make judgment. Peripheral cues are used to bypass logical reasoning and often invoke quick responses that are not well thought out.

The Elaboration Likelihood Model (ELM) and Heuristic Systematic Model (HSM) are very similar. Firstly, in the description of the cognitive evaluation process in dual modes as illustrated in Figure 4. Secondly, they both assume that people have a desire to hold onto what they judge to be the correct attitudes or judgment for a given scenario. The correctness could be determined by their evaluation of the arguments presented in the message but also their reliance on certain persuasive cues. Thirdly, both suggest that engagement in the higher cognitive effort (central/systematic) is driven by the processing of issue-relevant arguments. They also both agree that long-lasting attitudes and behaviour changes are affected by this higher cognitive effort.

Figure 4 illustrates a scenario where an individual is presented with a deceptive phishing web site and is required to make a judgement regarding its credibility. The individual is likely to engage in two possible cognitive evaluation processes during the evaluation phase depending on their motivation or ability. If the individual does not engage with the web site at all, there will be 'no evaluation'. If the individual is not motivated to evaluate the web site or does not have the necessary ability to evaluate it, for example, due to lack of skills, they may accept the deceptive signals in the phishing web site based on "heuristic" or "peripheral" evaluation. However, if they are both motivated and are able to evaluate the phishing web site, they would engage in what is termed as "systematic" or "central" evaluation.

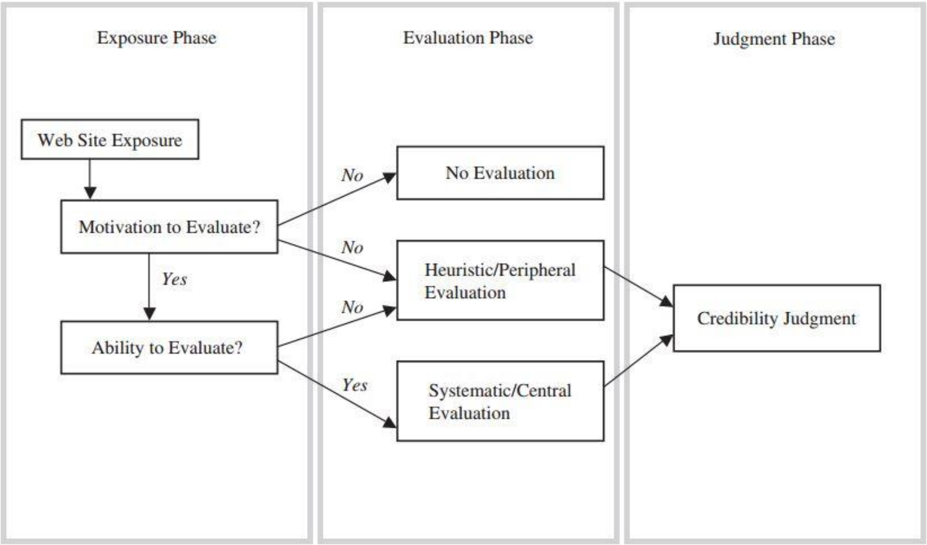


Figure 4: ELM and HSM Dual Processing Modes (Metzger, 2007)

Although ELM and HSM are largely similar, they have a few important differences as pointed out by the model developers (Eagly & Chaiken, 1993; Petty, 1994; Petty & Wegener, 1998). HSM posits that heuristic rules are knowledge structures that are kept in memory and accessed by an individual when they are evaluating a persuasive communication. In addition, HSM presents the concept of the “sufficiency threshold” whereby an individual only engages in evaluating a message until the sufficiency threshold is reached. When some initial heuristic processing does not meet the threshold then systematic processing is engaged. In contrast, ELM recognizes heuristic processing as just one of a number of possible peripheral route processes. In ELM there is a trade-off (negative relationship) between central and peripheral processing thereby giving a distinction for underlying attitude-forming processes as opposed to HSM in which both modes augment each other.

Due to these differences, and also the considered view that ELM evaluates a multi-dimensional space of the source, message, recipient and contextual factors (Petty & Wegener, 1999); it is proposed that ELM be used in this research. Additionally, ELM has been explored more widely in information system research; such as studies by: Wang et al. (2012) and Vishwanath et al. (2011) on factors that lead to phishing susceptibility; Angst & Agarwal (2009) on the acceptance of Electronic Health Record systems; Workman (2007, 2008b) on phishing and pretext social engineering; LaRose et al. (2008) on improving users’ online security behaviour; Bhattacherjee & Sanford (2006) on accepting new information technologies; and Johnson et al. (2001) on in the detection of financial fraud. This is in contrast to fewer information systems studies that have used HSM such as the study by Luo et al. (2013) on factors that lead to successful phishing.

The Elaboration Likelihood Model is illustrated in Figure 5. The flow diagram illustrates two routes of cognitive processing; the central route and the peripheral route. The central route leads to central attitude changes that are more enduring and predictive of long-lasting behaviour while the peripheral route leads to peripheral attitude shifts that are temporal and cannot be relied upon to predict long-term behaviour.

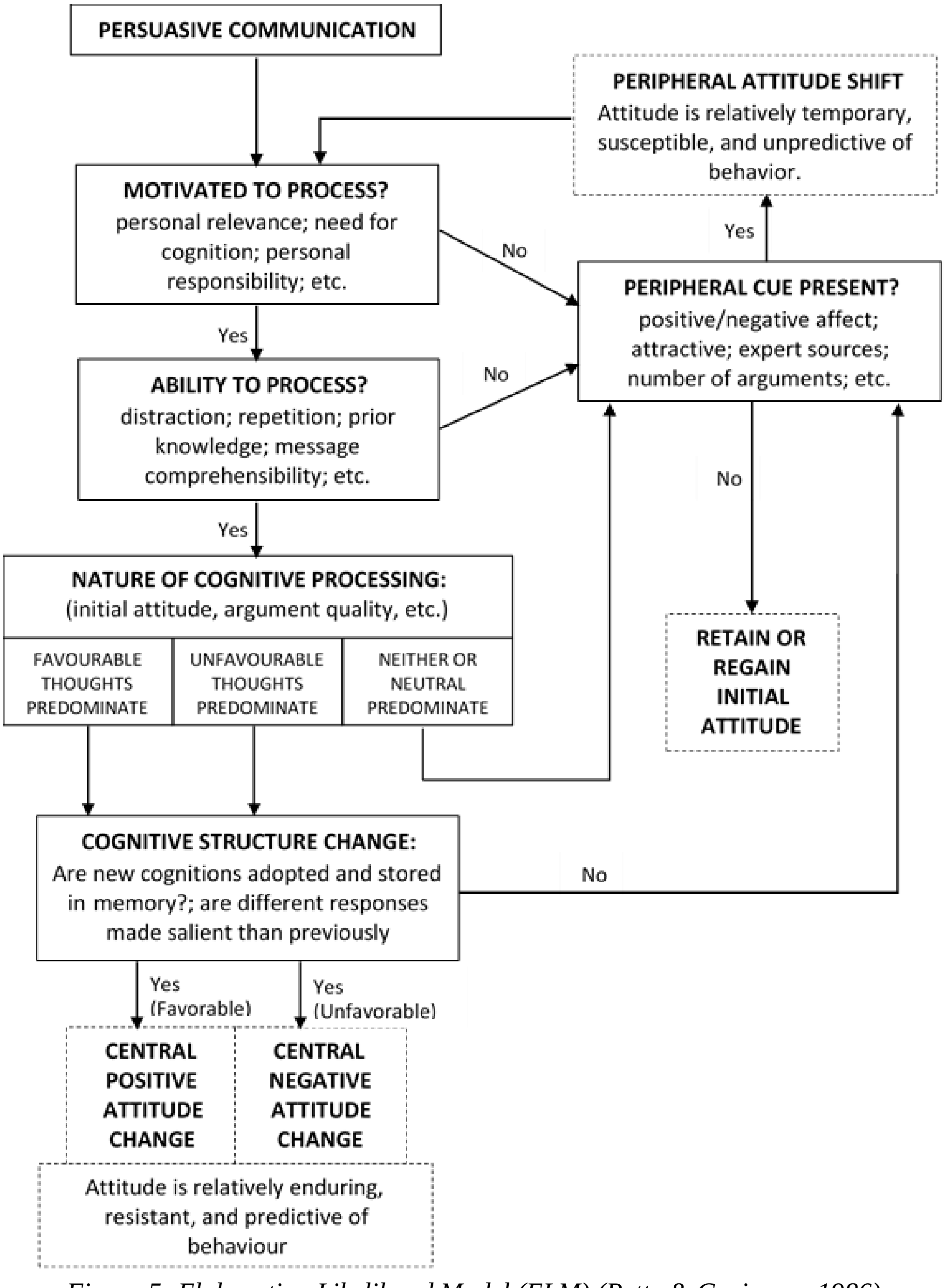

PERSUASIVE COMMUNICATION
PERIPHERAL ATTITUDE SHIFT
Attitude is relatively temporary, susceptible, and unpredictive of behavior.
MOTIVATED TO PROCESS?
personal relevance; need for cognition; personal responsibility; etc.
No
Yes
PERIPHERAL CUE PRESENT?
positive/negative affect; attractive; expert sources; number of arguments; etc.
Yes
ABILITY TO PROCESS?
distraction; repetition; prior knowledge; message comprehensibility; etc.
No
No
Yes
NATURE OF COGNITIVE PROCESSING:
(initial attitude, argument quality, etc.)
FAVOURABLE THOUGHTS PREDOMINATE
UNFAVOURABLE THOUGHTS PREDOMINATE
NEITHER OR NEUTRAL PREDOMINATE
RETAIN OR REGAIN INITIAL ATTITUDE
COGNITIVE STRUCTURE CHANGE:
Are new cognitions adopted and stored in memory?; are different responses made salient than previously
No
Yes (Favorable)
Yes (Unfavorable)
CENTRAL POSITIVE ATTITUDE CHANGE
CENTRAL NEGATIVE ATTITUDE CHANGE
Attitude is relatively enduring, resistant, and predictive of behaviour

Figure 5: Elaboration Likelihood Model (ELM) (Petty & Cacioppo, 1986)

Figure 5 illustrates the following steps in the ELM processing flow:

- An individual receives persuasive communication.
- If the individual is not motivated to process the persuasive communication, they take up the peripheral processing route. However, if the individual is motivated to process the persuasive communication (for example, because it is relevant to them), they proceed to the next step of the central processing route.
- If the individual is not able to process the persuasive communication (for example, because of distractions), they also take on the peripheral processing route. However, if the individual is able to process the persuasive communication they continue to the next step of the central processing route.
- When individuals are on the peripheral processing route, they will examine subjective criteria of the persuasive communication (for example, if they can identify themselves with the source, or if the look and feel of the message is credible). If convinced they will experience a peripheral attitude shift and subsequently take an action. If not, they will retain their initial attitude regarding the matter.
- Conversely, when individuals take on the central processing route because they are sufficiently motivated and able to process the persuasive message, they will examine objective criteria (for example, the issue-relevant arguments presented in the communication) and weigh their thoughts.
- If their thoughts are more favorable towards the persuasive message, they will experience a central positive attitude change and subsequently positive behaviour change. However, if their thoughts are more unfavorable, they will have a central negative attitude change and subsequently negative behaviour change. If they feel they cannot rely on their thoughts to come up with a convincing conclusion, they will retain their initial attitude and behaviour regarding the matter.

A key construct in ELM is “Elaboration” which describes the mental effort an individual engages when evaluating persuasive communication. High elaboration means the individual is engaged in high levels of objective information processing and is associated with the central route of information processing. Low elaboration means the individual is engaged in low levels of biased information processing which tends toward subjective reasoning associated with the peripheral route. The key factor for an individual’s ability to detect phishing attacks is their level of elaboration of the phishing scenario. In fact, Vishwanath et al. (2011) demonstrated that successful phishing attempts are mostly characterized by low elaboration. Luo et al. (2013) point out that an attacker's aim is to generate phishing communication that discourages objective, systematic processing but encourages attention to deceptive peripheral cues that result in quick and incorrect decisions.

The deceptive peripheral cues identified in previous studies on phishing susceptibility include: spelling and grammar; professional look and feel; genre conformity; security padlock icons; endorsements; spoofed or falsified source credibility; hiding the deception behind text or images; pretexting; urgency and time pressure (Downs et al., 2006; Jakobsson et al., 2007; Kumaraguru, Sheng, et al., 2007; Luo et al., 2013; Vishwanath et al., 2011).

It is also important to emphasize that for an insider's thought processing to be successful in identifying deception, they should be knowledgeable on both the threat domain and the deception cues. Various studies have demonstrated this (Dhamija et al., 2006; Dodge, Carver, & Ferguson, 2007; Downs et al., 2006, 2007; Jakobsson et al., 2007; Vishwanath et al., 2011) and have shown that the more knowledgeable insiders are on the threat domain and detection cues, the less likely they are to succumb to unintentional threats. These studies have also pointed out that this knowledge could be obtained from training and awareness activities but also from an insider's past exposure to a similar threat.

The concept of "Elaboration" is similar to the concept of "Activation" that is advanced by the Theory of Deception. However, unlike the Theory of Deception, ELM provides differentiation of the information processed into the categories of issue-relevant arguments and peripheral cues and examines the effect that paying attention to these different components has in detection of deception. This enables us to elucidate the different components of deception and examine their effect on the detection of deception.

Another key contribution that distinguishes ELM from other theories is that it seeks to understand what would make the individual (1) motivated to process the persuasive communication presented to them and also (2) what would interfere with their ability to process it objectively. It posits that people will be motivated to process persuasive communication if they feel involved in or responsible for the matter presented. The more a person is motivated to process, the more likely they are to have higher levels of elaboration. On the other hand, their ability to process is hindered by factors such as distraction, emotions or pressure. The lower a person's ability to process the more likely they will have a lower level of elaboration.

Cialdini's (2001) principles of influence and persuasion, namely; authority, scarcity, liking and similarity, reciprocation, commitment and social proof provide a complementary resource to enrich the understanding of factors that affect an insider's motivation and ability to process persuasive communication. Similarly, work by Bezuidenhout, Mouton, & Venter (2010) and (Peltier, 2006) identify other factors to consider, such as; strong affect/emotions, overloading, deceptive relationship and diffusion of responsibility.

The Interpersonal Deception Theory, Theory of Deception and the Elaboration Likelihood Model only focus on the persuasive signal, interpretation and response concepts relating to cognitive factors but do not adequately examine other individual, contextual factors and organizational factors that may affect people's susceptibility to Unintentional Insider Threats. It is therefore important to bring in other theories that can help address this. One theory that has been proposed in previous studies by Workman (2007, 2008a) and Workman et al. (2008) is the Protection Motivation Theory.

2.2.4 Protection Motivation Theory

The Protection Motivation Theory (PMT) by Rogers (1975, 1983) helps predict people's responses when faced with a threat. The Protection Motivation Theory was primarily used in the health sciences to understand how to motivate people to take up healthy lifestyles. It examines how an emotion, such as fear, can lead to behaviour change. However, it has also been used in numerous studies in the information systems discipline such as the study by Tsai et al. (2016) on enacting online safety behaviours; Bojmaeh (2015) on end-user information systems security behaviour; Anderson & Agarwal (2010) on home computer users' security-related behaviours; Vance et al. (2012) on motivating information systems security compliance; Johnston & Warkentin (2010) on the use of fear appeals to influence end users to comply with recommended information security actions; Herath & Rao (2009b) on information systems security policy compliance; Lee & Larsen (2009) and LaRose et al. (2008) on improving users' online security behaviour; Workman et al. (2008) on omission of information security features; Pahnila et al. (2007) on information systems security policy compliance; Woon et al. (2005) on securing home wireless networks; and Liang & Xue (2009, 2010) on technology threat avoidance.

When a fear appeal is communicated and received by an individual, the Protection Motivation Theory describes two mediating cognitive processes of "Threat Appraisal" and "Coping Appraisal" that the individual uses to determine their attitude and behaviour response.

The original Protection Motivation Theory model by Rogers (1975) decomposed the Threat Appraisal process into: (1) perceived severity of the threat; and (2) perceived susceptibility to the threat – also termed as perceived vulnerability. The Coping Appraisal process consisted of a construct that measures a person's assessment of the effectiveness of the recommended responses - also termed the *response efficacy*. Rogers (1983) later revised the PMT and borrowed from Bandura (1977) to include an additional construct called *self-efficacy* to the Coping Appraisal process; which is a person's evaluation of their ability to execute the recommended response. In addition, considerations were also made regarding the *perceived cost* and *perceived benefit* of recommended responses. This approach of examining the cost and benefits has been shown to affect behavioural response in various studies (Herath & Rao, 2009b; Lee & Larsen, 2009; Weinstein, 1993; Workman et al., 2008). If an individual's assessment is that the costs (physical effort, monetary expenditure, time usage or even cognitive exertion) outweigh the benefits (threat avoidance); then they are unlikely to take up the safeguarding measure.

These constructs that make up the Protection Motivation Theory are illustrated in Figure 6. The Protection Motivation Theory constructs are very useful in determining a person's behavioural response in unintentional insider threat scenarios. In many cases, an insider's response is informed by their perception of the threat (Threat Appraisal) and their ability to act in a secure manner (Coping Appraisal). Studies have shown that these constructs influenced by an insider's knowledge of the threat and their skills in executing a recommended response (Kumaraguru et al., 2009, 2008; Kumaraguru, Rhee, Acquisti, et al., 2007; Kumaraguru, Rhee, Sheng, et al., 2007; Kumaraguru, Sheng, et al., 2007; Sheng et al., 2007).

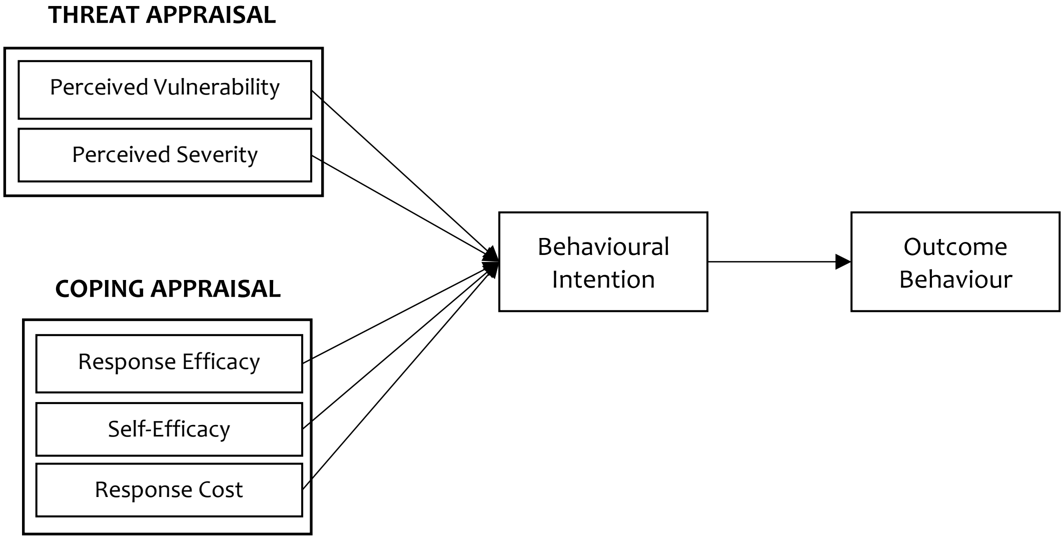


Figure 6: Protection Motivation Theory (Rogers, 1975, 1983)

Previous work that uses the Protection Motivation Theory in the context of unintentional insider threats has been done by Workman (2007, 2008a) and Workman et al. (2008). However, these studies only examined "threat appraisal" constructs leaving out the "coping appraisal" constructs. In contrast, this research will examine both "threat appraisal" and “coping appraisal" constructs.

2.2.5 Technology Threat Avoidance Theory

Liang & Xue (2009, 2010) present the Technology Threat Avoidance Theory (TTAT) which outlines similar constructs to those presented by the Protection Motivation Theory; namely, the ‘Threat Appraisal’ and ‘Coping Appraisal’ constructs.

However, it differs from the Protection Motivation Theory by proposing additional constructs. These are: Perceived Threat, Perceived Avoidability, Avoidance Motivation and Avoidance Behaviour as the behavioural outcome. In addition, Risk Tolerance and Social Influence constructs are presented in the model as illustrated in Figure 7. These additional constructs present new insights in the study of threat avoidance behaviours associated with unintentional insider threats as opposed to acceptance behaviours that are often associated with intentional insider threats (Liang & Xue, 2009).

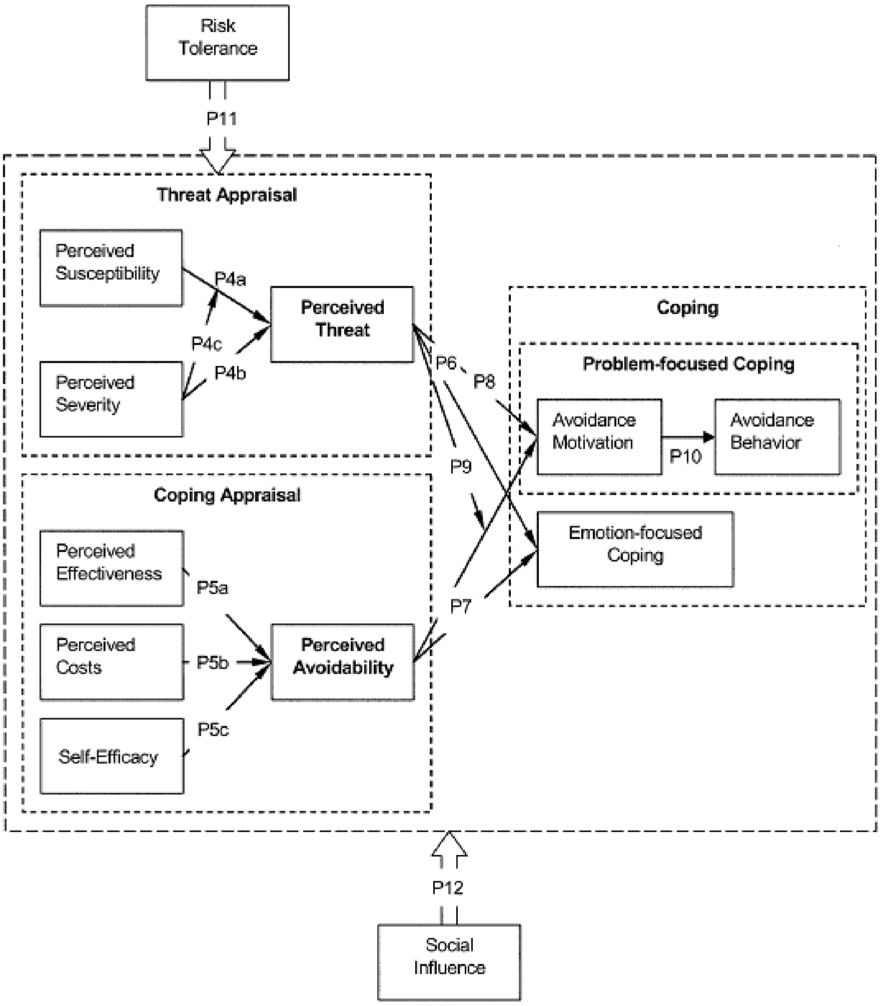


Figure 7: Technology Threat Avoidance Model (Liang & Xue, 2009)

Unlike the Protection Motivation Theory, the Technology Threat Avoidance Theory tries to address the cognitive evaluation process by distinguishing ‘Problem-focused Coping’ from ‘Emotion-focused Coping’. However, it does not do so to the level of depth and clarity that Heuristic Systematic Model (HSM) and Elaboration Likelihood Model (ELM) do.

The Technology Threat Avoidance Theory also does not explore persuasion and deception constructs. Attackers targeting insiders often use persuasive and deceptive tactics in order to manipulate insiders to compromise the security of their systems (Luo et al., 2013; Vishwanath et al., 2011; Wang et al., 2012). Therefore, these constructs are important and need to be considered.

The Technology Threat Avoidance Theory has not been tested extensively and is not supported by many empirical studies (Yasin, Fatima, Liu, Yasin, & Wang, 2019).

Liang & Xue (2010) conducted the first empirical study grounded on the Technology Threat Avoidance Theory to study how a convenience sample of 152 business students at a university use anti-spyware to protect themselves from spyware threats. However, their study excluded four constructs, namely: perceived avoidability, emotion-focused coping, risk tolerance and social influence. This study therefore did not fully test the Technology Threat Avoidance Theory. In addition, the study population and sampling techniques did not allow the findings to be generalizable.

2.3 Empirical Foundation

Various studies have advanced the understanding of antecedent factors that lead to Unintentional Insider Threats (UIT). Homoliak et al. (2018) did a review of the best ranked and most cited literature in the insider threat domain. They came up with a list of 322 works from 1980 to 2018 but filtered out 108 of them from their analysis due to their exclusion criteria that disregarded studies that did not examine insider threat as their main subject or that had presented the same study across multiple papers. In their analysis, they provided a taxonomy that is useful in categorizing the existing literature in a structured way. This taxonomy organizes existing work in what they term a 5W1H methodology that categorizes studies based on who, what, where, when, why and how. Their taxonomy does not particularly delve into articulating a theoretical or conceptual framework to understanding unintentional insider threats. They however identify psychological, social and criminal theories that some authors have advanced in relation to insider threats.

Ophoff et al. (2014) conducted a similar review of literature on insider threats. They had a wide survey scope and did not zero in on unintentional insider threats. They found 90 unique articles from the top 50 ranked Information Systems Journals using the search term “Insider Threat”. They then classified these 90 articles into 6 categories and 13 sub-categories. The categories and sub-categories were: insider threat overview (definition, case studies or examples); insider threat behaviour (unintentional, intentional, motives); theoretical perspectives (application of existing theory, advancement); insider threat mitigation (non-technical, technical detection, technical prevention); insider threat management (governance, regulatory); and miscellaneous. The articles found first dated back to 1997 up to 2013 with most articles being from 2008 and 2009. They found a total of 52 articles with some form of theoretical background in the research. However, they did not provide a list of articles that they had reviewed and neither did they provide the statistics for the Unintentional Insider Threat sub-category. It is likely their search term “insider threat” was too broad to capture studies that focus on specific types of unintentional insider threat; such as, social engineering.

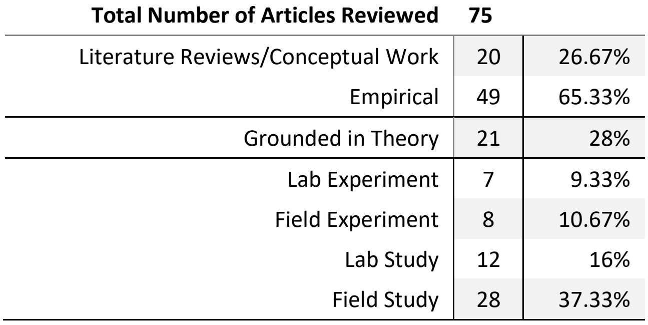
Tetri & Vuorinen (2013) took a similar approach in their review of the existing body of knowledge relating to insider threats. However, they specifically focused on unintentional insider threats and narrowed down to a list of 40 papers. In their analysis, only 5 had some analysis of empirical data and only 2 had explicit theoretical underpinnings. This clearly shows that the UIT area is under-researched, firstly from a theoretical view and also from an empirical standpoint; a view supported by others (CERT, 2013; Greitzer et al., 2014; Ophoff et al., 2014). However, Tetri & Vuorinen's (2013) literature review focused on social engineering and therefore did not consider studies exploring other attack vectors.

2.3.1 Overview of Empirical Studies

This thesis presents a review of 75 articles that specifically study the unintentional insider threat phenomenon. An outline of these studies is provided in Table 1. Of these 75 studies that were reviewed, 49 (65.33%) are based on empirical findings and 20 (26.67%) are literature reviews of previous work or conceptual presentations of ideas. Only 21 of these 75 studies (28%) are grounded in theory confirming a deficiency in the use of theory to study the unintentional insider threat as pointed out in previous work (Luo et al., 2013; Tetri & Vuorinen, 2013; Wang et al., 2012; Workman, 2007).

Of these 75 studies that were reviewed seven were lab experiments (9.33%), eight were field experiments (10.67%), twelve were lab studies (16%) and twenty-eight were field studies (37.33%). It should be noted that field studies and field experiments are preferred over laboratory research because they have a higher ecological validity especially when respondents are examined in a real-world natural environment (Huber, Kowalski, Nohlberg, & Tjoa, 2009; Kumaraguru et al., 2009, 2008; Workman, 2007, 2008a). Lab studies and experiment are subject to bias and Hawthorne effects because participants know they are being studied and often modify their behaviour. If participants are not alerted about the ongoing investigation, results are able to objectively reflect their normal behaviours (Downs et al., 2007; Vishwanath et al., 2011). In addition, if participant selection is done randomly and is reflective of the typical population then the results are also highly generalizable (Bowen, Salem, Hershkop, Keromytis, & Stolfo, 2009).

Table 1: Summary Statistics of the Literature Reviewed



Total Number of Articles Reviewed	75	
Literature Reviews/Conceptual Work	20	26.67%
Empirical	49	65.33%
Grounded in Theory	21	28%
Lab Experiment	7	9.33%
Field Experiment	8	10.67%
Lab Study	12	16%
Field Study	28	37.33%

2.3.2 Constructs Identified from Empirical Studies

The existing body of work suggests various constructs that affect a user’s susceptibility to unintentional insider threats. These are discussed hereafter.

1. *Ability to Process*

Work by Workman (2007, 2008a, 2008b); Vishwanath et al. (2011) and Luo et al. (2013) examined a user’s ability to process a threat scenario in order to determine their susceptibility to unintentional insider threats. Workman (2007, 2008a, 2008b) explored the effect various emotions such as fear, trust, sense of commitment, obedience and reactance have on a person’s ability to process a threat scenario. They conducted a field study that staged various social engineering attacks using phishing emails, websites, phone calls and employing various deceptive techniques. They targeted 850 participants of a government-regulated entity that had experienced serious security breaches. The studies show that people who have higher levels of fear, trust, commitment, obedience and reactance are more susceptible to unintentional insider threats.

It should be noted that these constructs are presented by Cialdini (2001) who explored deception techniques that attackers commonly use to manipulate insiders. Cialdini's Principles of Influence and Persuasion include: authority, scarcity, liking and similarity, reciprocation, commitment and social proof. Workman (2007, 2008a, 2008b) studies tested all these constructs and found them to have an influence on susceptibility to unintentional insider threats. The studied examined authority using obedience and fear, scarcity using reactance, liking using trust, reciprocation using normative commitment, commitment using continuance commitment and social proof using affective commitment.

In their study, Vishwanath et al. (2011) used the Theory of Deception to examine the effect email load has on a person's ability to process a threat scenario. They explain that if a person receives a high volume of emails they tend to pay less attention and effort in processing specific elements of each email. This may increase the likelihood of a person responding to phishing emails without paying much attention to them and therefore posing an unintentional insider threat. Their field study results were not able to prove the effect email load has on attention. Their study was however able to show that to support the hypothesis that email load had a significant effect on the likelihood to respond to phishing emails.

Luo et al. (2013) based his work on Heuristic Systematic Model (HSM) by Chaiken (1980). They examined how a user's ability to process an attack is affected by 'Time Pressure'. They explained that time pressure for immediate action influences a person's ability to process a threat scenario. If a person feels pressed for time or is under pressure to take immediate action, they are more likely to fall victim to Unintentional Insider Threats. In their field study, they conducted a spear phishing campaign that targeted faculty and staff of an US public university. The study did not incorporate time pressure strongly. Therefore, the study was unable to conclusively test its effect.

The Elaboration Likelihood Model (ELM) by Petty & Cacioppo (1986) presented 'Distraction' as an additional factor that influences a person's ability to process. To the best of our knowledge, this construct has not been empirically studied in the unintentional insider threat literature. They explain that the more a person is distracted, the more effort they have to extend to try and process a message. Distractions lower a persons' ability to process. They posit that distractions should enhance persuasion for messages that do not have strong logical arguments. Therefore, deceptive messages are more likely to employ distractions as a way to enhance persuasion.

2. *Motivation to Process*

Studies by Vishwanath et al. (2011), Wang et al. (2012) and Luo et al. (2013) examined a person's susceptibility to unintentional insider threats based on the their motivation to process a threat scenario. Vishwanath et al. (2011) examined the Involvement construct derived from the Elaboration Likelihood Model (ELM) by Petty & Cacioppo (1986). They defined involvement as the perceived relevance a message or scenario had to a person. They hypothesized that a person is more likely to process a phishing message if it seems relevant to them. The more relevant the message is, the more likely they will get involved with it and the more motivated they will be to process it. The results of their field study showed that involvement had a significant influence on attention given to urgency cues. In addition, involvement had a significant influence on the processing of the deceptive message.

In their study, Wang et al. (2012) presented a research model based on the Theory of Deception. They examined the Involvement construct based on the hypothesis that a person will expend more cognitive effort to process a message based on the degree to which they perceive it to be pertinent to them. Their field study surveyed people who had been targeted by a spear phishing attack at an US public University. Results of their analysis showed that involvement significantly increases the cognitive effort in processing phishing emails.

Luo et al. (2013) examined how the 'Need for Cognition' construct affected a person's motivation to process an attack scenario. They explain that a need for cognition is the desire a person has to comprehend and structure the scenario presented to them. People with a high need for cognition have a greater ability to process a threat scenario and are therefore less likely to fall victim. They found that faculty and staff who had high levels of education or who paid keen attention to detail had higher need for cognition. However, their study did not measure the levels of need for cognition and neither did it make a direct link between need for cognition and actual susceptibility to the attack.

The Elaboration Likelihood Model (ELM) by Petty & Cacioppo (1986) presents 'Personal Responsibility' as an additional factor that influences the motivation to process. To the best of our knowledge, this construct has not been studied in the unintentional insider threat literature. They explain that the more a person associates a sense of personal responsibility regarding a communication, they more likely they will engage with the communication. Spear phishing emails take advantage of this factor because the more personalized a message is, the more it will connect with a person's sense of responsibility to act.

It is important to point out that these studies did not examine the direct relationship 'involvement', 'need for cognition' or 'personal responsibility' have on the outcome unintentional insider threat behaviour. The effects these constructs have on susceptibility to unintentional insider threats were examined through indirect effects.

3. *Attack Factors*

Tetri & Vuorinen (2013) put in a strong case for a multi-dimensional understanding of the unintentional insider threat. They explain that a lot of previous work has focused on the weaknesses of the victim but have not addressed other perspectives. This research classifies two constructs (Quality of Argument and Persuasive Cues) from the Elaboration Likelihood Model (ELM) by Petty & Cacioppo (1986) as attack factors. These two constructs examine the techniques employed by an attacker in order to make their attack successful.

The effect persuasive cues have on susceptibility to unintentional insider threats has been examined empirically in studies by: Grazioli (2004), Jakobsson (2005), Karakasiliotis, Furnell, & Papadaki (2006), Workman (2007, 2008a, 2008b), Huber et al. (2009), Vishwanath et al. (2011), Wang et al. (2012) and Luo et al. (2013). However, the effect that quality of argument has on susceptibility to unintentional insider threats has been studied less.

Grazioli (2004) examined six persuasive cues in their field study at a US university. They examined the effect that (1) forged third-party assurance seals, (2) forged news clips and quotes from professional magazines, (3) 'too-good-to-be-true' warranty statements, (4) picture of a store's physical location, (5) website sales volumes and (6) testimonials have on the ability of people to detect deception. They gathered responses from 80 students pursuing an MBA in Information Systems who were deemed to be both business and IT savvy. They found that only 15% of them could detect deception while 35% incorrectly assessed the deceptive site as not deceptive.

22.5% of the students were undecided. This showed that majority of the IT savvy respondents could not detect deception despite their specialized training. Through hypothesis testing they discovered that the respondents who were able to detect deception paid less attention to the persuasive cues. Those who were deceived paid more attention to the trust-based persuasive cues.

Jakobsson (2005) presented a graph model and used it to describe various cases of context-aware phishing attacks. A context-aware phishing attack is one which is highly relevant and believable based on prevailing circumstances. A good example is when an attacker cuts a user’s telephone lines, then waits for them to call the telephone company to report disconnection, then the attacker knocks on their door and introduces themselves as a repairman from the phone company who will fix their network if let in. They carried out a simple field study involving 25 users on eBay and found that 60% of the participants could be victims of a staged context-aware phishing attack. They also conducted a second survey study and the results showed that only one of the respondents was suspicious of the communication they received. This meant as high as 96% of the survey respondents could have been victims of the staged context-aware phishing attack.

Karakasiliotis, Furnell, & Papadaki (2006) conducted a field study involving 179 participants interacting with a web-based survey questionnaire that presented 20 emails. Eleven (11) of the emails were illegitimate while 9 were legitimate. Participants were asked to classify each of the emails as either ‘legitimate’, ‘illegitimate’ or ‘I don’t know’. The results showed that 32% of the responses incorrectly classified the emails and 26% indicated that they did not know. The study also analyzed 1,653 feedback comments from the respondents in order to understand their judgement criteria. They found that 40 participants relied on visual cues such as logos, banners and fonts that are often imitated in phishing emails. Nineteen (19) participants said they relied on presence of language mistakes such as typographical and grammatical errors to make their decision. Eighteen (18) participants said that the presence of their names in the phishing emails gave them legitimacy while sixty-seven (67) said that the presence of their account numbers gave the messages legitimacy. One short coming of this study was that the 20 emails were not interactive and legitimacy could only be judged from examining the email as presented. Other methods like visiting live sites, using tools, search engines, domain information or browser indicators could not be used to judge legitimacy. The experiment also primed the respondents to look for deception, which would not happen in naturalistic, real-life attacks.

Workman (2007, 2008a, 2008b) incorporated Cialdini's (2001) Principles of Influence and Persuasion in their staged field study and experiment. These can be seen as techniques incorporated in an attack to make them more successful. The study involved using student-actors who were trained and coached to conduct the phishing and pretext deceptions.

Jakobsson et al. (2007) noted that a high degree of attack personalization increased trust. Jagatic et al. (2007) demonstrated that by sending spear phishing emails that were customized from social network profiles, attack success rates were as high as 72% as opposed to generalized phishing emails at 16% success. Dhamija et al. (2006) found that well designed phishing sites fooled 90% of the participants.

Huber et al. (2009) used an Automated Social Engineering (ASE) Bot to stage social engineering attacks on the Facebook social networking site. The ASE bot crawled the social networking site to identify victims based on profiles that had weak security and privacy settings. The bot was able to remain active for 3 days and to send more than 100 chat messages with targeted victims within that period. The ASE bot incorporated Cialdini's (2001) Principles of Influence and Persuasion to increase success of the attacks.

Vishwanath et al. (2011) examined the effect of persuasive cues based on the attention a person gave to the (1) email source (sender’s name, email and reply-to address); (2) grammar and spelling (typographical errors, content and grammar in title and body); (3) urgency cues (warnings and statements indicating urgency of a time-bound nature) and (4) subject line (seen by recipient before email is opened). These were specific techniques incorporated in the phishing email. Results of their study showed that all these persuasive cues had a significant effect on an individual’s likelihood to respond to phishing emails. Attention to the email source and grammar and spelling had a negative effect while attention to urgency cues and subject line had a positive effect on the likelihood to respond to phishing emails.

Wang et al. (2012) examined the effect attention given to persuasive cues (termed as visceral triggers in their study) had on cognitive processing and also the likelihood to respond to the phishing message. They explained that visceral triggers are designed to decrease rational thought processing and decision making; with the result of a person being a victim of an attack. The visceral trigger that the study focused on was the stressing of urgency of response to a phishing email. The study analyzed 321 completed responses from undergraduate students of a state university in the US. Results of their study showed that attention to visceral triggers had a significant negative impact on cognitive processing but a significant positive impact on the likelihood to respond to a targeted spear phishing email.

Unlike the previous studies that only examined the effect of persuasive cues, Luo et al. (2013) also examined the effect quality of argument had on susceptibility to unintentional insider threats. They also examined a construct termed ‘Pretexting’ that refers to a technique which attackers use to make their messages more believable by using existing contexts. The pretext that they used in their study was based on rumored budget cuts and delayed budget talks that were a current topic of interest at the university. Their field study targeted 105 students and staff at a public university in the US. They used the term heuristic cues to refer to persuasive cues in line with the Heuristic Systematic Model (HSM) by Chaiken (1980). The persuasive cues examined were: (1) source credibility and (2) genre conformity. Source credibility involved creating an email address that would appear genuine to the recipients. Genre conformity involved replicating the look and feel of legitimate communication that was sent in the university. This involved imitating the logos, fonts, phrases and overall layout of the phishing email and website to make them look and feel familiar. The study analyzed the factors in a qualitative way and was not able to conclusively test the hypothesis presented. However, the exploratory analysis of the responses showed that participants were more likely to be victimized by phishing messages that have a high argument quality. In addition, participants were more likely to be victimized by phishing messages that have source credibility and genre conformity as persuasive cues. The pretexting component designed into the attack was also analyzed. Analysis also showed that phishing attacks with pretexting are more likely to succeed in victimizing targeted insiders.

It can be argued that pretexting as demonstrated by Luo et al. (2013) builds on the argument quality to make the deceptive message more believable. This is similar to the concept of context-aware phishing attacks presented by Jakobsson (2005). This allows for these constructs to be examined within the theoretical foundation provided by the Elaboration Likelihood Model (ELM) by Petty & Cacioppo (1986).

4. *Elaboration*

Studies by Vishwanath et al. (2011) and Wang et al. (2012) empirically examine the elaboration construct in the context of unintentional insider threats. Elaboration is coined from the Elaboration Likelihood Model (ELM) by Petty & Cacioppo (1986) which refers to the extent to which a person expends cognitive effort in processing or thinking about the issue-relevant arguments in a persuasive message. Elaboration is measured on a continuum that ranges from high to low elaboration. High elaboration is associated with the expense of considerable cognitive effort to objectively examine the issue-relevant arguments in a persuasive message. Conversely, low elaboration is associated with less cognitive effort in cognitive processing of the issue-relevant arguments in a message. Petty & Cacioppo (1986) argue that high elaboration is driven by the effect of three antecedent constructs: Quality of Argument, Motivation to Process and Ability to Process. They also point out that low elaboration is driven by the effect of persuasive cues.

Vishwanath et al. (2011) examine the effect cognitive processing of four persuasive cues (email source, grammar and spelling, urgency cues and subject line) have on the level of elaboration. The results of their study show that only urgency cues have a significant effect on elaboration but the other three do not. Their results further demonstrate that urgency cues lower the level of elaboration. This study also examines the effect that elaboration has on the likelihood to respond to phishing emails. They find that elaboration has a negative relationship on the likelihood to respond to phishing but this effect was not significant. This means that the higher the elaboration, the less likely a person is to respond to a phishing message.

Wang et al. (2012) refer to the elaboration construct as ‘Cognitive Effort’ in their study. They examine the effect cognitive effort has on reducing the likelihood to respond to targeted phishing attacks. The results of their study show that cognitive effort in processing phishing emails does reduce the likelihood of responding to targeted phishing attacks. However, the effect was not significant.

5. *Knowledge*

Various studies examine the effect the knowledge construct has on susceptibility to unintentional insider threats (Aytes & Connolly, 2004; Dhamija et al., 2006; Downs et al., 2006, 2007; Fogg et al., 2001; Friedman, Hurley, Howe, Felten, & Nissenbaum, 2002; Garera, Provos, Chew, & Rubin, 2007; Grazioli, 2004; Jakobsson & Ratkiewicz, 2006; Jakobsson et al., 2007; Karakasiliotis et al., 2006; Sheng et al., 2010; Tsow & Jakobsson, 2007; Vishwanath et al., 2011; Wang et al., 2012).

Knowledge is examined from various perspectives, namely knowledge relating to: the threat domain, detection cues and determinants of trust. Knowledge on the threat domain involves an understanding of the Unintentional Insider Threat and how it is perpetrated. Knowledge on detection cues underscores an awareness of cues that identify deceptive messages. Knowledge on determinants of trust pertains to an understanding of indicators that mark legitimate and trustworthy messages.

Fogg et al. (2001) in their field study, with data from 1410 online survey participants from Europe and the US, examined respondent's knowledge on various elements that are used to determine the credibility of websites. Fifty-one (51) exploratory elements were eventually grouped into 7 factors through factor analysis: (1) real-world feel; (2) ease of use; (3) expertise; (4) trustworthiness; (5) tailoring; (6) commercial implications and (7) amateurism. The 'Real-World Feel' measurement considered whether: a website provided quick response to customer enquiries, if the site had listed a physical address, phone number, email address or even photos of the organization members. The 'Ease of Use' scale considered whether: the site had a search functionality, looked professionally designed, was arranged in a way that made sense, did not take long to download and was not difficult to navigate. The 'Expertise' scale examined if: the site was run by an organization that was well respected outside of the internet, it listed credentials of authors of its content, its articles gave citations and references, it had few but detailed articles, it declared to be the official site for specific content, had ratings or reviews for its content and if it displayed awards it had won. The 'Trustworthiness' scale considered if: the site was linked to by sites that were believable, it stated its content policy, it linked to outside materials and sources, it provided links to competitor sites, it was recommended by a friend, it represented a non-profit organization, it listed well-known corporate customers or its URL ended with .org. The ‘Tailoring’ scale examined if: the site sent emails to confirm transactions, it selected news stories based on user preferences, it recognized previous visits and it required registration or login. The ‘Commercial Implications’ scale considered if: the site advertised on radio or billboards, it had advertisements that matched the content, it was designed for e-commerce, it did not have a commercial purpose, it did not require paid subscription, it had no adds, it did not have automatic popup windows with ads and it did not make it hard to distinguish advertisements from content. The ‘Amateurism’ scale considered if the site had been updated, it offered information in more than one language, it was not less than 5 pages, it was not hosted by third parties, its domain name matched the company name, it had no typographical errors, it did not become unexpectedly unavailable, it had no links that did not work, it did not link to sites that were not credible and if it was rarely updated with new content.

The Fogg et al. (2001) study focused on determinants of trust. However, subsequent studies argue that some of the factors listed are not sound criteria for judging credibility. Phishing messages and websites can fake a lot of the criteria in an aim to deceive insiders to trusting the communication.

The Friedman et al. (2002) study was also on determinants of trust. It engaged seventy-two (72) individuals from three different communities through interviews to understand user conceptions of web security. They found that the participants relied primarily on six characteristics to determine secure web connections: (1) use of HTTPS Protocol; (2) presence of a lock or key icon; (3) point in transaction, for example if the main-page is secured; (4) type of information, for example, request for social security numbers or passwords; (5) type of website, for example, bank websites are expected to be secure and (6) general distrust, for example, users generally thinking no website is secure.

Grazioli (2004) used the Theory of Deception to underscore the importance of three knowledge areas, namely: threat domain, detection cues and determinants of trust. Individuals need knowledge on the threat domain to enable them to know how attackers may use deceptive tactics. He added that individuals detect deception by being able to identify cues of deception through picking anomalies in the communication they are processing. He pointed out that according to the Theory of Deception, individuals needed superior knowledge on the cues to look for and how to interpret them when evaluating normal verses deceptive communication in order to detect deception. He emphasized that successful and unsuccessful detectors of deception could see the same cues but may arrive at different conclusions based on their knowledge and skill at interpreting the cues. He also classified cues into two categories; assurance cues and trust cues. The results of data analysis showed that the individuals who were successful at detecting deception relied on a correct assessment of third-party and legally binding cues such as seals and warranties. However, individuals who were unsuccessful at detecting deception used unverifiable cues used to build trust but could be fake, such as customer testimonials.

Karakasiliotis et al. (2006) had the aim of investigating how user's knowledge on deception ploys and techniques affects their susceptibility to Unintentional Insider Threats. They surveyed 179 participants and used over 1,653 feedback comments on the criteria they used to judge the credibility of websites. Fifty-two (52) participants said that they used technical cues relating to the URL shown in a message to determine if it was credible. Forty (40) participants said they used visual factors such as fonts, logos, banners, trademark and copyright symbols to determine credibility. Other participants said they relied on content characteristics such as language mistakes, personalization, offers and forceful instructions to judge credibility. Their key finding was that despite using these different judgement criteria, participants often arrived at incorrect decisions regarding the legitimacy of content. The interpretation of technical cues found on the URL and within emails was often wrong.

Dhamija et al. (2006) performed a cognitive walkthrough of about 200 sample attacks on an Anti-Phishing Working Group archive. They identified three strategies that are used to deceive people and these were: (1) lack of knowledge; (2) visual deception and (3) lack of attention. They categorized the lack of knowledge dimension into two areas. The first was a lack of computer systems knowledge relating to the identification of fraudulent URLs and email spoofing techniques (detection cues). The second category was a lack of knowledge on security indicators (determinants of trust). Their lab study engaged 22 participants in assessing 20 fully-functioning websites to determine if they were legitimate or deceptive. Their study results showed that well designed phishing websites fooled 90% of the users. Majority of the users (59%) did not know how to use security indicators (determinants of trust) to assess credibility of websites. These users had the lowest score for correctly identified sites because they looked at website content and did not assess the URL, status bar or security indicators such as SSL encryption. Only 31.8% of the users knew how to examine the padlock icon in the browser, SSL encryption indicators or encryption certificates when judging websites. This knowledge was vital for correctly classifying legitimate and fraudulent websites. Worse still, 68% of the participants admitted to disregard security warnings concerning fraudulent certificates and proceeded to access phishing websites. Many (86%) said that they had never checked a certificate before. In addition, 31% of the participants said that they had never heard of the term phishing indicating a lack of knowledge in the threat domain.

Jakobsson & Ratkiewicz (2006) designed a phishing study that intentionally incorporated cues that could be detected by knowledgeable users but would deceive naïve users. They used URLs that appeared questionable, modified transaction queries to spoof identity and created malicious links. These intentional indicators of phishing were designed into the experiment to test the knowledge of users on their use in detecting phishing. The results of their study showed that 11% of users who encountered a message with obvious phishing indicators (detection cues) would still be susceptible to the attack because they did not have knowledge on how to evaluate various security indicators (determinants of trust).

Downs et al. (2006) conducted a lab study involving 20 non-expert users. The users were engaged in a role-play exercise to examine 8 emails in the inbox of a fictitious user called Pat Jones. They were asked to explain the actions they would take for each email and the reasons for their decisions. Thereafter, the participants were interviewed to understand their knowledge on trust indicators and awareness on various security cues and a mostly qualitative analysis of their responses was done. Eighty five percent (85%) of the participants said that they had seen padlock images on websites but very few could explain what the padlock meant or how to interpret it correctly. They could not tell if the padlock was used as a deceptive cue or as a true sign of security. Thirty five percent (35%) of the participants noticed a difference between HTTP and

HTTPS indicators in the websites. However, awareness of these very general security cues did not translate into meaningful behaviour change.

Tsow & Jakobsson (2007) conducted a field study that engaged 435 participants through an online questionnaire survey to examine screenshots of six emails and six webpages. The screenshots were carefully designed to incorporate deceptive cues and determinants of trust. The participants were asked to rate the level of 'phishiness' or authenticity using a 5-point Likert scale. Results of the study indicated that users did not know how to validate URLs or to determine legitimacy of HTTPs trust indicators.

Jakobsson et al. (2007) conducted a lab study with 17 participants being shown 26 samples that had a mixture of both phishing and authentic content. The samples were designed with endorsement logos, domain names, IP addresses, padlocks in various positions, spelling and grammatical irregularities, HTTP and HTTPS links and personalized content to assess participant knowledge on deception cues and determinants of trust. The participants were to rate the 'phishiness' or authenticity of each sample on a 5-pont Likert scale and to verbally describe the reasons for their decision. A mainly qualitative analysis of the responses was done. Results showed that users ignored many trust indicators in favor of spelling and grammar. Presence of padlock icons drew attention but did not improve trust. Endorsements by third parties were only effective if the brand was recognized.

Downs et al. (2007) conducted a field study that engaged 232 participants in an online questionnaire survey. The first section showed participants images of 5 emails and asked them to play the role of Pat Jones and to determine the actions they would take. The respondents were then given 4 URLs and asked to determine what they could tell about the website just by examining the URL alone. Next, the participants were shown 4 padlock icons with one positioned within the browser frame but three others not within the browser frame. They were asked to interpret the meaning of these padlock icons. The next section of the survey asked the participants to choose the best definition for 4 computer terms in order to test their knowledge on the threat domain. The results of their study showed that participants who correctly answered the question testing their knowledge on phishing were less likely to fall for phishing. This indicated that the knowledge on the phishing threat domain showed a familiarity with the concept that significantly protected them from the risk. Participants who had the correct knowledge on the meaning of padlock icons as a determinant of trust also were significantly less likely to fall for phishing attacks. In addition, participants who could analyze hyperlink URLs correctly to identify deception cues were significantly less likely to fall for phishing attacks.

Garera et al. (2007) used a lab study to test a framework that can be used to detect phishing attacks using heuristics associated with URLs. They proposed 18 features, organized in 4 groups, to identify phishing URLs and sites. They tested their framework using several million URLs collected from Google’s Safe Browsing toolbar. Results of their study showed that their classifier had 97.3% accuracy in detecting phishing URLs. Although their work focused on implementing an automated classifier, user knowledge of the 18 features of phishing URLs can significantly help them avoid phishing attacks.

Kumaraguru, Sheng, et al. (2007) found that 42% of the participants used an assessment of design and content as the strategy for identifying phishing attacks. They point out that this is a bad strategy because most phishing content is copied and carefully designed to imitate legitimate content. Additionally, many users do not look for concrete trust determinants. Only 3% of the participants used security indicators to identify phishing attempts. In fact, Kumaraguru, Rhee, Acquisti, et al. (2007) observed that 80% of the participants were unable to use a simple technique that examines the address behind hyperlinks using the mouse-over technique. Similarly, Dhamija et al. (2006) found that 59% of users never looked for ‘HTTPS’ in the address, 68% never paid attention to the padlock security icon, 77% did not notice SSL indicators given by browsers in the address bar, 86% had never examined a security certificate and, worse still, 68% disregarded pop-up warnings.

Sheng et al. (2010) administered an online survey to 1001 participants that engaged them in a role-play task to evaluate a set of emails where six were phishing, five were legitimate, two were spam messages and one email containing possible malware. Participants were asked to specify how they would handle each email. The study specifically examined the effect knowledge has on phishing susceptibility using two techniques. Firstly, participants were asked to complete a knowledge test to assess their general knowledge on phishing before the role play task. Secondly the participants were given anti-phishing training and then asked to complete a second role play task.

Analysis was done to see if the anti-phishing training had an effect on reducing phishing susceptibility. Results of their study showed that students who scored highly on the threat domain knowledge test were significantly less likely to fall for phishing. For every standard deviation increase in the knowledge test score, there was 3.6% less likelihood to fall for phishing. Results on the analysis of the effect of anti-phishing education on phishing susceptibility showed that participants who had prior anti-phishing education significantly predicted phishing susceptibility. Participants who had no prior anti-phishing education fell for 60% of the phishing websites. In addition, the anti-phishing training designed into the study gave a 40% reduction in participants falling for phishing between the first role play task and the second. Each of the four types of anti-phishing training reduced the susceptibility to phishing. There was a control group that did not receive any anti-phishing training. Results showed that there was no significant improvement in their ability to detect phishing between the first role play task and the second. This study demonstrated the impact knowledge has in reducing susceptibility to attacks.

Vishwanath et al. (2011) did an analysis of 161 university email users who had been targeted by two phishing campaigns. They examined the indirect effect that prior domain-specific knowledge had on phishing susceptibility through the elaboration construct. Their knowledge construct considered general knowledge about emails, email-based scams and also about university emails. Results showed that domain-specific knowledge had a significant effect on elaboration in one of the cases. This knowledge on the threat domain gave users more confidence and accuracy in decision making during elaboration.

Wang et al. (2012) conducted a study grounded on the Theory of Deception that also examined the effect the knowledge construct has on susceptibility to unintentional insider threats. They conducted a field study that collected 267 valid responses from an online questionnaire. The results of their study showed that knowledge weakens the effect of persuasive cues while strengthening the attention to deception indicators; which both significantly reduce the likelihood of falling for targeted phishing attacks. In addition, their results showed a marginally significant direct effect between knowledge and reduction in phishing susceptibility.

6. Threat Appraisal

The Threat Appraisal construct from the Protection Motivation Theory (PMT) by Rogers (1975, 1983) has been examined in a number of studies to understand its effect on susceptibility to unintentional insider threats (Arachchilage & Love, 2013; Aytes & Connolly, 2004; Downs et al., 2006, 2007; Liang & Xue, 2009, 2010; Luo et al., 2013; Workman, 2007, 2008a; Workman et al., 2008). The Threat Appraisal construct can be decomposed into perceived severity and perceived vulnerability.

Aytes & Connolly (2004) conducted a field study that collected 167 questionnaire responses and analyzed the effect perceived vulnerability and perceived severity had on user security behaviours. Three specific outcome user behaviours relating to passwords, emails and data backup were examined. Results of their study showed that users recognize that there are significant negative consequences (perceived severity) to insecure computer behaviours but they still engage in insecure behaviours. In addition, their work showed that users believe the possibilities of these negative consequences happening to them (perceived vulnerability) is fairly low, therefore explaining why perceived severity had little effect.

Downs et al. (2006, 2007) examined the effect that ratings of negative consequences (perceived severity) had on phishing susceptibility. They did not explicitly examine the Threat Appraisal construct but their interpretation of ratings of negative consequences can be associated with perceived severity. Downs et al. (2006) results of their lab study involving 20 non-expert users showed that the users who engaged in more online activities and those who had been victims of online fraud rated negative consequences highly. This perceived severity did not make the users less likely to fall victim to phishing attacks. Downs et al. (2007) examined 232 online questionnaire responses and found that perceived severity of consequences had a significant effect in the unwillingness of users to interact with legitimate sites (false positives). This meant that the users generally feared interacting even with trustworthy sites.

Workman (2007, 2008a) and Workman et al. (2008) examined the Threat Appraisal construct by examining the Threat Severity (Perceived Severity) and Threat Vulnerability (Perceived Vulnerability) constructs. They found that those with a lower perceived severity are more susceptible to social engineering attacks. Similarly, those with a lower perceived vulnerability are more susceptible to social engineering attacks. The Workman et al. (2008) study tested the effect that three treatments (warnings on policy violations, ethics training and social engineering training) had on social engineering. The results of the study showed that none of the treatments had an effect on perceived severity or vulnerability. This meant that people still perceived social engineering to have a major threat severity and susceptibility regardless of the treatment intervention applied.

Liang & Xue (2009) incorporated both Threat Appraisal constructs in their Technology Threat Avoidance Theory (TTAT). These constructs were empirically evaluated in the Liang & Xue (2010) field study using 152 online questionnaire responses. Results showed that there was a third mediating variable which they termed 'Perceived Threat'. Perceived Threat fully mediated the effect that Perceived Susceptibility (vulnerability) and Perceived Severity had on Threat Avoidance Motivation. Subsequently perceived threat significantly influenced the outcome Avoidance Behaviour. However, they did not provide results showing the direct effect of Perceived Vulnerability and Perceived Severity on the overall outcome Avoidance Behaviour.

Arachchilage & Love (2013) also examined the Perceived Threat, Perceived Susceptibility (vulnerability) and Perceived Severity constructs from the Technology Threat Avoidance Theory (TTAT) by Liang & Xue (2009). Their field study collected 151 responses from undergraduate students using a questionnaire survey. Results of their analysis showed that perceived threat is determined significantly by the perceived vulnerability and perceived severity variables. Just as in the Liang & Xue (2010) study, they found that perceived threat fully mediated the effect of perceived vulnerability and perceived severity on threat avoidance motivation. Unlike the Liang & Xue (2010) study, they found that perceived threat is also influenced by the interaction between perceived severity and perceived vulnerability. Perceived Threat subsequently had a significant effect on the outcome Avoidance Behaviour. However, they did not examine the direct effect of perceived vulnerability and perceived severity on the outcome avoidance behaviour.

Luo et al. (2013) study gave results of a qualitative analysis of a spear phishing attack that targeted 105 faculty and students at a public US university. They did not specifically examine the threat appraisal construct but examined the effect that perceived damage had on phishing susceptibility. Their description of perceived damage is very similar to perceived severity. Results showed that phishing attacks that seemed to cause less damage (asking for a small amount of seemingly innocuous information) were more likely to victimize targeted users. Their analysis showed that users felt that they did not have much to lose and therefore responded to phishing requests. They also pointed out that attackers use this strategy to increase the success of their attacks. The attacker would use tactics that are careful not to raise concern of targeted users so that they can comply with their phishing requests.

7. *Perceived Threat*

The Technology Threat Avoidance Theory (TTAT) by Liang & Xue (2009) presents a fairly new construct to the unintentional insider threat research named Perceived Threat. This construct is empirically texted by Liang & Xue (2010) and Arachchilage & Love (2013) and is shown to fully mediate the effect that the threat appraisal constructs (perceived vulnerability and perceived severity) have on Threat Avoidance. The Perceived Threat construct is also significantly influenced by the interaction effect between perceived vulnerability and perceived severity. This interaction effect means that if either perceived vulnerability or perceived severity is not present (has a score of 0) then the effect of the other construct is nullified. It is important to incorporate this new construct into this Unintentional Insider Threat study.

8. *Coping Appraisal*

Just like the Threat Appraisal construct, the Coping Appraisal construct is presented by the Protection Motivation Theory (PMT) by Rogers (1975, 1983). It is composed of response efficacy, self-efficacy and response cost sub-constructs. These constructs have been examined in various Unintentional Insider Threat studies (Arachchilage & Love, 2013; Aytes & Connolly, 2004; Bojmaeh, 2015; Liang & Xue, 2009, 2010; Vishwanath et al., 2011; Workman et al., 2008).

Workman et al. (2008) examine the effect of the self-efficacy, response efficacy and response cost constructs in their study. Their study examined 612 responses from a large technology-based services company. Their data was from both self-reported questionnaire responses but also from directly observed behaviours. Three outcome behaviours were examined: (1) password updates and protection; (2) security and anti-virus updates and (3) system backups. The results of their analysis showed that all three constructs, self-efficacy, response efficacy and response cost-benefit had significant effects on the outcome security behaviours. Their results showed that people with high perceived self-efficacy, high perceived response-efficacy and high perceived response benefit over cost were less likely to omit taking security precautions.

The Technology Threat Avoidance Theory (TTAT) by Liang & Xue (2009) also incorporates the Coping Appraisal constructs from the Protection Motivation Theory (PMT). These constructs are empirically tested by Liang & Xue (2010) and Arachchilage & Love (2013) in their studies. Results show that self-efficacy and safeguard effectiveness (response efficacy) have a significant positive effect on avoidance motivation. In contrast, response cost has a significant negative effect on avoidance motivation. These studies do not examine the direct effects these three variables have on the outcome Avoidance behaviour.

9. *Organizational factors*

Various Information Security best-practice frameworks such as ISO27000 series, COBIT and NIST Special Publications in Information Security, advocate for organizations to use a mixture of policy, technology and people-based security measures for the effective protection of information systems (ISACA, 2012; ISO, 2013; Nieles, Dempsey, & Pillitteri, 2017).

Various empirical studies have also examined the effect these different security measures have on the mitigation of Unintentional Insider Threats (Aytes & Connolly, 2004; Bojmaeh, 2015; Downs et al., 2006, 2007; Egelman et al., 2008; Huber et al., 2009; Kumaraguru et al., 2009, 2008, 2008; Kumaraguru, Rhee, Acquisti, et al., 2007; Kumaraguru, Rhee, Sheng, et al., 2007; Kumaraguru, Sheng, et al., 2007; Sheng et al., 2010, 2007). However, it should be noted that more has been done on the use of technology and people security measures than on the use of policy.

Aytes & Connolly (2004) examined 167 responses from undergraduate students regarding their security behaviours. Although a vast majority of the respondents rated themselves as knowledgeable regarding security behaviours, 47% of them reported as never having received any information or training on computer security matters. Of those who said they had received some information, only 19% of them said they had received it through formal training or education. Majority of them, 52%, cited friends and co-workers as the source of their information and 42% said they had gained it from personal experience. The assessment of their actual security behaviours reflected this gap. Results showed that 49% of the respondents engaged in risky security behaviour occasionally and 28% did so frequently or all the time.

Downs et al. (2007) collected survey responses from 232 participants at the Carnegie Mellon University in USA. In order to profiling the respondents, they were asked what computer security measures they had implemented in the past to protect themselves. Ninety three percent (93%) stated that they had installed anti-virus software on their machines and 79% stated that they had adjusted their browser security settings.

Kumaraguru, Sheng, et al. (2007) conducted a lab experiment that had two groups: a control group that did not receive any anti-phishing training and a second that received anti-phishing training. Each group comprised of 14 participants. Each participant was asked to review a set of 20 emails and to state whether they were legitimate or phishing emails. A pretest was conducted with 10 emails (5 phishing and 5 legitimate), then a training intervention was given to the treatment group and finally a post test was done with a different set of 10 emails (5 phishing and 5 legitimate). Results showed a significant decrease in false negatives after training showing that training of users is an effective way of preventing phishing. However, the training group also recorded an increase in the false positives; meaning users also incorrectly marked legitimate sites as phishing sites. The increase in false positives was not significant but it reduced the overall correctness rate.

Kumaraguru, Rhee, Acquisti, et al. (2007) conducted a lab experiment to study the effect of training on phishing susceptibility. They set up three treatment groups with each group consisting of 10 participants who are taken through different types of training. One group was given standard email security notices sent by e-commerce organizations to alert their customers about phishing. The second group was given text and graphics training and the third group was given a comic strip-based training. Results showed that standard text-based notices sent through email were not effective in teaching people about phishing. The comic-strip training intervention which used little text and a story based graphical theme was most effective in reducing phishing susceptibility.

Kumaraguru, Rhee, Sheng, et al. (2007) conducted another lab experiment to study the effect anti-phishing education has on phishing susceptibility. They set up three groups: one control group that did not receive any training, one treatment group that received non-embedded training and a final treatment group that received embedded training. They define embedded training as training that is delivered just after a user takes an insecure action when interacting with staged phishing emails. The study was conducted in two sessions that were 7 days apart in order to see if the participants retained the knowledge they had acquired over the 7 days. Results showed that the embedded training group significantly improved in their ability to detect phishing messages after training than the other two groups. Training that is embedded in a user's normal day-to-day activities is more effective than training that is detached from their activities. However, there was no significant difference between the performance of the non-embedded training group after training and the control group in identifying phishing messages. This showed that embedded training increased the ability of users to detect phishing but the non-embedded training did not. In order to test retention, the participants were tested again after 7 days. Results showed that even after 7 days, participants in the embedded training group still performed significantly better than their counterparts in the non-embedded training and control groups. After the 7 days, only 7% of the participants in the non-embedded training group were able to correctly identify a phishing attack compared to 64% of participants in the embedded training group. In addition, users in the embedded training group were able to transfer their knowledge onto different phishing scenarios as compared to their counterparts in the non-embedded training and control groups.

Sheng et al. (2007) conducted another study to evaluate the effect of a game called Anti-phishing Phil that teaches people not to fall for phishing. They designed a lab experiment with three treatment conditions: one that reviewed existing training materials, another which went through an anti-phishing tutorial and a final group that used the Anti-phishing Phil game. All participants were asked to evaluate 10 websites, then complete a 15-minute training task and subsequently evaluate another different set of 10 websites. Results showed that all three treatment conditions that went through training significantly improved the participant's ability to identify phishing websites. However, the Anti-phishing Phil game condition registered the best performance in reducing false negatives and false positives and its participants reported the greatest confidence in their decision making.

The study by Egelman et al. (2008) examined the effect technology had on phishing susceptibility through the use of browser warnings. In the lab experiment, 60 participants were assigned to one of four groups: one 20-member group received Firefox browser warnings, the second 20-member group received active Internet Explorer warnings, the third 10-member group received passive Internet Explorer warnings and the fourth 10-member control group received no warnings at all. Results showed that 89% of the participants were susceptible to phishing attacks. Active browser warnings (that interrupt users and force them to attend to the warnings) were found to be significantly more effective than passive warnings. In addition, there was no significant difference between the control group and the group that received passive browser warnings; indicating that passive warnings were not effective in preventing phishing attacks. Results also showed a significant relationship between recognizing browser warnings and ignoring them because of seeing them multiple times (termed as habituation). This means that users tend to dismiss warnings that they have seen multiple times. Also, there was a significant correlation between trusting warnings and obeying them. When users trusted the warnings they received, they heeded them.

Kumaraguru et al. (2008) conducted a field experiment to test the effect of embedded training in a real-world environment. They engaged 311 participants who worked in a Portuguese company in a field experiment that had 3 groups: one control group where 111 participants did not receive any training, a second where 100 participants received generic training and a third group where 100 participants received embedded training when they interacted with a spear phishing message. Results of their study showed that both training conditions significantly reduced participants' susceptibility to phishing after training. However, there was no significant difference between the generic training materials and the embedded training. A key limitation in this study was that participants in the different groups discussed the study among themselves and exchanged ideas because they were in close proximity with each other.

This had the effect of contaminating the results especially of the participants assigned to the control condition.

Kumaraguru et al. (2009) conducted a field experiment over 28 days involving 515 participants at Carnegie Mellon University who were divided into three groups: a control group that received no training, a treatment group that received one training and another treatment group that received two trainings. Results of the study showed that participants in the control group showed no significant changes in their ability to identify phishing messages over the study period. Participants who either received one training or two training interventions performed significantly better in avoiding phishing attempts after training. In addition, participants who were trained twice performed significantly better than those who only received one training. In addition, the training did not reduce the participants' willingness to interact with legitimate emails (false positives). The one-train group demonstrated that users retained knowledge from training even after 28 days. The two-train group demonstrated that repeated training reinforced knowledge and reduced susceptibility even further.

Huber et al. (2009) conducted a field experiment using an Automated Social Engineering (ASE) Bot to crawl Facebook social networking site and preform social engineering attacks. It should be noted that the ASE bot was able to use automated features to harvest information from people's accounts because of weak security and privacy settings. This highlights the need to tighten technical security controls for online accounts and social media profiles to prevent information gathering and subsequent social engineering attacks.

Sheng et al. (2010) administered an online survey to 1001 participants and later assigned them to one of five groups; a control group that received no training, a second experiment group that received popular web-based training, a third experiment group that used the Anti-phishing Phil game to learn, a fourth experiment group that used the PhishGuru cartoon and a final experiment group that received both Anti-phishing Phil and PhishGuru training. Prior to training the participants indicated that they would have clicked on 52% of the phishing links and 47% would provide information on phishing websites. This indicated that 90% of the participants who would click the link would go ahead to submit information on phishing sites. After training the number of participants who fell for phishing websites reduced by 34% to 44%. This was regardless of whether the training used conventional materials or specialized materials developed along learning science principles. There was no significant improvement in the ability to detect phishing in the control group. In addition, the Anti-phishing Phil and PhishGuru training did not decrease the willingness of participants to interact with legitimate websites (false positive). Only the conventional training materials made the participants avoid even the legitimate websites.

Bojmaeh (2015) was able to demonstrate that the use of security technology (antivirus and spam filters) had a significant positive impact on information security behaviour. These various studies have shown that user training (particularly embedded training) and use of security technologies (particularly antivirus, spam filters and browser tools) are able to significantly reduce user susceptibility to Unintentional Insider Threats.

10. Threat Avoidance

Liang & Xue (2009) in the Technology Threat Avoidance Theory (TTAT) present a new construct in the study of Unintentional Insider Threats called Threat Avoidance. This construct is tested in subsequent studies by Liang & Xue (2010) and Arachchilage & Love (2013). Results of the empirical studies show that Threat Avoidance Motivation has a significant positive effect on the outcome Threat Avoidance Behaviour. The avoidance motivation is very similar to the behavioural intention construct in the Protection Motivation Theory by Rogers (1975, 1983).

11. Demographic Factors

A number of studies have also examined the effect of various demographic factors on susceptibility to Unintentional Insider Threats. These variables have mainly been tested as demographic variables in order to give focus to the factors being examined in the research model. They include: Gender, Age, Education, Department, Years on the Internet, Hours on the Internet, Computer Skill, Email Load, Online Service Usage, Prior Victimization and Risk Propensity.

i. Gender

Gender was examined in studies by Fogg et al. (2001); Friedman et al. (2002); Aytes & Connolly (2004); Dhamija et al. (2006); Tsow & Jakobsson (2007); Jakobsson et al. (2007); Jagatic et al. (2007); Downs et al. (2007); Kumaraguru, Sheng, et al. (2007); Kumaraguru, Rhee, Acquisti, et al. (2007); Kumaraguru, Rhee, Sheng, et al. (2007); Sheng et al. (2007); Workman (2007, 2008b); Egelman et al. (2008); Kumaraguru et al. (2008); Kumaraguru et al. (2009); Huber et al. (2009); Liang & Xue (2010); Sheng et al. (2010); Wang et al. (2012); Arachchilage & Love (2013).

Many studies have shown that gender has no significant influence on susceptibility to Unintentional Insider Threats (Arachchilage & Love, 2013; Dhamija et al., 2006; Egelman et al., 2008; Kumaraguru et al., 2009, 2008; Kumaraguru, Rhee, Acquisti, et al., 2007; Kumaraguru, Rhee, Sheng, et al., 2007; Kumaraguru, Sheng, et al., 2007; Liang & Xue, 2010; Sheng et al., 2007; Wang et al., 2012; Workman, 2007, 2008a, 2008b).

However, Jagatic et al. (2007) and Sheng et al. (2010) showed that women were more likely to click on phishing links and submit information on phishing websites than men. Fogg et al. (2001) also showed that men assigned less credibility to the websites presented in their study than women. However, Sheng et al. (2010) performed further analysis and found that gender differences could be attributed to differences in technical knowledge and training. A mediation analysis using technical knowledge and technical training as the mediators, showed that women were more susceptible to phishing because they had less technical knowledge and training than men.

Huber et al. (2009) conducted a study using an Automated Social Engineering (ASE) Bot to target Facebook users. They proposed a study scenario where the bot would be setup to take the profile of a 22-year-old single female student from Great Britain with an attractive profile picture in order to target single male victims with a malicious phishing link. This scenario shows that success of social engineering attacks can be increased by using sexually attractive gender identities. This was also confirmed by Jagatic et al. (2007) in a similar study that used techniques to crawl on social networking sites to target victims. They found that the susceptibility to phishing increased if an attack was staged with a profile of someone of the opposite gender. This was true for both males and females.

ii. Age

Various studies also examined the effect age has on susceptibility to Unintentional Insider Threats. Some studies found that age had no significant influence on susceptibility to Unintentional Insider Threats (Arachchilage & Love, 2013; Dhamija et al., 2006; Egelman et al., 2008; Kumaraguru, Sheng, et al., 2007; Liang & Xue, 2010; Sheng et al., 2007). However, other studies found that age had a significant influence on susceptibility to Unintentional Insider Threats (Fogg et al., 2001; Jagatic et al., 2007; Kumaraguru et al., 2009; Sheng et al., 2010; Wang et al., 2012; Workman, 2007, 2008b).

Sheng et al. (2010) and Kumaraguru et al. (2009) found that participants in the 18-25 years age group were more susceptible to phishing than older age groups. Sheng et al. (2010) did further analysis using a multiple-mediator model incorporating four mediators (prior exposure to training, education level, years on the internet, financial risk investing). The mediation analysis established that the reason for this was that this age group consisting of young people had a less exposure to training, lower level of education, few years on the internet and less averse to financial risks. Fogg et al. (2001) in their study of what makes websites credible found that respondents under the age of 27 were more critical of websites that seemed amateurish (having typing and grammatical errors) than those older than 37 who rated websites with markers of expertise and trustworthiness higher.

iii. Education

Studies have also examined the effect education has on susceptibility to Unintentional Insider Threats. Dhamija et al. (2006); Sheng et al. (2007) and Kumaraguru, Sheng, et al. (2007) found that education had no significant correlation to susceptibility to phishing. However, studies by Dodge et al. (2007); Workman (2007); Workman (2008b); Sheng et al. (2010) found that the level of education had a significant effect on susceptibility to Unintentional Insider Threats.

Dodge et al. (2007) studied phishing susceptibility on the student population at the United States Military Academy. Level of education was based on the 4 class years (freshmen, sophomore, juniors and seniors) at the academy. Results of their study showed that those in upper classes were less susceptible to phishing and were more likely to report phishing attacks than those in lower classes. The reason for this was that students who had been at the academy longer had received annual information security awareness training.

Sheng et al. (2010) found that the level of education had the most significant impact on phishing susceptibility compared to all other demographic factors. It also explained why female participants and young participants within the 18-25 years age group were significantly more susceptible to phishing. They conducted mediation analysis and found that the reason why these groups were significantly more susceptible was because they had a lower level of education, less technical knowledge and training.

Fogg et al. (2001) examined the effect the level of education had on analyzing the credibility of websites. They found that participants who had completed graduate education assigned more credibility to websites that displayed markers of trustworthiness.

iv. Department Specialization

Some studies have examined the effect of department specialization on susceptibility to Unintentional Insider Threats. Jagatic et al. (2007) found that students who were pursuing technology degrees in computer science, informatics and cognitive science were the least likely to fall for phishing attacks. In fact, none of the science students in the control group fell for the phishing attack. Kumaraguru et al. (2009) also confirmed this when they found that computer savvy people were less likely to fall for phishing. Sheng et al. (2010) further corroborated these findings and showed that for every standard deviation increase in technical knowledge, participants of the study fell for 3.6% fewer phishing attacks. For this reason studies by Downs et al. (2006, 2007); Sheng et al. (2007); Jakobsson et al. (2007); Kumaraguru, Rhee, Sheng, et al. (2007) deliberately excluded participants who could be thought of as computer experts in order to make their results more generalizable to a wider population.

Kumaraguru et al. (2009) found a significant difference between participants from the academic and the administrative departments in susceptibility to phishing. A close examination as to why participants in the academic department were more vulnerable to phishing showed that it was because of the large number of students who were assigned to the academic department.

v. Years on the Internet

A few studies have examined the effect that number of years spent by users on the internet has on susceptibility to phishing. Kumaraguru, Sheng, et al. (2007) collected data on how many years participants had used the internet and correlated it with their susceptibility to phishing. They found no significant correlation. In addition Fogg et al. (2001) measured user experience with the internet using three questions that measured; years they have used the internet, hours they spend online and the number of purchases they make online. Just like Kumaraguru, Sheng, et al. (2007), the study by Fogg et al. (2001) did not find any significant effect.

However, Sheng et al. (2007) was able to show a statistically significant correlation between years on the internet and the ability to correctly identify phishing attempts. In addition, the study by Sheng et al. (2010) was able to explain why the younger 18-25 age group was significantly more susceptible to phishing by using the measure of years on the internet as a mediator.

vi. Hours on the Internet

Some studies have examined the effect that hours on the internet has on susceptibility to Unintentional Insider Threats. Studies by Fogg et al. (2001) and Arachchilage & Love (2013) used the term 'internet experience' that was measured by the number of hours that participants spent online. Sheng et al. (2007) also measured the number of hours spent online per week as a demographic characteristic. However Dhamija et al. (2006) measured the number of hours spent using the computer per week. Kumaraguru, Sheng, et al. (2007), Sheng et al. (2007), Liang & Xue (2010) and Arachchilage & Love (2013) found no statistically significant influence of hours on the internet on user susceptibility to phishing.

vii. Computer Skill

Studies by Jagatic et al. (2007), Kumaraguru et al. (2009) and Sheng et al. (2010) showed that people who were computer and technology savvy were less likely to fall for phishing. Computer savviness was determined by the degree that a person had in the study by Jagatic et al. (2007). However, it was judged on a self-rated scale in the study by Sheng et al. (2010). Studies by Downs et al. (2006), Kumaraguru, Rhee, Acquisti, et al. (2007), Kumaraguru, Rhee, Sheng, et al. (2007), Sheng et al. (2007) and

Egelman et al. (2008) used screening questions to identify computer-savvy users. Users were considered computer savvy if they had either changed browser preference settings, created websites or helped someone fix a computer problem. Kumaraguru, Rhee, Sheng, et al. (2007) deemed users who knew what phishing was as computer savvy.

Majority of these studies excluded computer savvy users from their studies because they believed these users would bias their results and make them less generalizable. However, Egelman et al. (2008) chose not to exclude them from their study because results of their pilot showed that they were just as likely to be susceptible to certain phishing attacks as their non-computer-savvy counterparts.

viii. Email Load

A number of studies have also examined the effect email load has on user susceptibility to Unintentional Insider Threats. Vishwanath et al. (2011) pointed out that this variable is often overlooked in research. They measured email load as the number or emails that an individual receives in a day. They explained that a large number of emails would be overwhelming therefore reducing the level of attention given to important elements of emails. Another consequence would be that a person is more likely to respond to phishing emails unconsciously just in order to reduce their load. The results of their hypothesis testing supported that a large volume of emails would reduce the level of attention paid but this was not to a significant level. However a large volume of email was found to significantly affect the likelihood to respond to phishing attacks. Kumaraguru, Rhee, Sheng, et al. (2007), Kumaraguru, Rhee, Acquisti, et al. (2007) and Sheng et al. (2010) collected demographic data on the average number of emails that participants received in a day but did not draw a relationship to phishing susceptibility in their analysis and results.

ix. Online Service Usage

Various studies collected demographic data on online shopping experiences but many of them did not use this data to draw a relationship with susceptibility to Unintentional Insider Threats. Egelman et al. (2008) collected demographic data on online shopping experience and Fogg et al. (2001) measured the number of purchases on the web. In the study by Downs et al. (2006), 95% of the participants indicated that they had purchased online and 70% had done online banking. Dhamija et al. (2006)

found that 82% of their participants regularly used online banking while 91% regularly shopped online. However, these studies did not analyze the relationship between the use of online services and susceptibility to phishing.

Downs et al. (2006) study was able to show that younger people were more significantly engaged in online activities than older people. In addition, Downs et al. (2007) found that participants who had experience with spoofed sites, eBay or PayPal, were less likely to click on phishing links.

x. Prior Victimization

In their study, Downs et al. (2006) reported that one-quarter of their study participants had been victims of fraud associated with the use of their credit card or social security number. One might expect that this group would be more wary of future attacks and less susceptible to phishing in the study. Surprisingly, results showed that they were marginally more likely to fall for the phishing in the study. The researchers attributed this to lack of knowledge; the same thing that could have made them victims in the first place. Therefore, this study showed that prior victimization did not guarantee a lower susceptibility to future attacks.

Workman (2008b) demographically profiled the respondents based on previous victimization to social engineering. They analyzed if these people were more likely to fall for their staged social engineering attack. The results showed that they were less susceptible to their staged social engineering attack but this finding was not statistically significant. Kumaraguru, Rhee, Acquisti, et al. (2007) noticed that one of the ten participants in an experiment condition group did not click on any hyperlinks and the reason they gave to the researchers was that they had been a victim of identity theft in the past.

xi. Risk Propensity

Sheng et al. (2010) showed that participants' risk aversion was a predictor to phishing susceptibility. Results of their study showed that more risk-averse participants were less likely to fall for phishing. In fact for each standard deviation increase in the risk perception measure, the participants fell for 2.8% less phishing. Downs et al. (2006) study was able to show that younger people were more significantly engaged in risky online activities than older people. The CERT (2013) foundational study explains that

Unintentional Insider Threat cases usually involve insiders who are more likely to take risks than the average individual. When risk averse individuals fall victim to Unintentional Insider Threats it is usually because of the influence of other factors such as stress, time pressure, workload or illness.

2.3.3 Summary of Studies

Sections 2.2 and 2.3 have discussed numerous studies that constitute the body of knowledge on Unintentional Insider Threats. An analysis of these studies has identified various factors that influence a user's susceptibility to Unintentional Insider Threats. In order to provide a good overview of these emergent factors despite the lengthy discussions, they are summarized in Table 2. Each study that examined a particular factor is listed by author, date and the country where the study was conducted. Subsequently, each study is annotated with the use of a ✓ symbol to indicate whether a particular factor was examined in the study. The factors are also organized based on higher-level constructs with an aim of determining the constructs that should be considered in this research.

In addition, 33 studies that are considered instrumental in making key contributions to the study of Unintentional Insider Threats are summarized in a different format in Table 3. Each study is listed by author and date in descending order starting from the most recent. In addition, a short description is given of each study to identify the theoretical foundation, research methodology, population, sampling and a critique is given of its key findings and limitations.

Table 2: Factors Extracted from Literature

	Constructs	Outcome Behaviour		Behavioural Intention	Organizational Factors			Coping Appraisal				Threat Appraisal		Knowledge			Elaboration		Attack Factors		Motivated to Process			Ability to Process			Demographic Factors											
	List of Relevant Empirical/Theoretical Studies: (Authors, Date) Country	**Secure**	**Insecure**	**Threat Avoidance**	**Policies**	**Technology Controls**	**Training & Awareness**	**Response Efficacy**	**Self-Efficacy**	**Response Cost**	**Threat Detection**	**Perceived Vulnerability**	**Perceived Severity**	**Threat Domain**	**Detection Cues**	**Determinants of Trust**	**Central Route**	**Peripheral Route**	**Quality of Argument**	**Persuasive Cues**	**Involvement**	**Need For Cognition**	**Responsible**	**Distraction**	**Emotions**	**Pressure**	**Gender**	**Age**	**Education Level**	**Role**	**Years on Internet**	**Hours on Internet**	**Computer Skill**	**E-mail Load**	**Email Responsiveness**	**Online Services Usage**	**Prior Victimization**	**Risk Propensity**
1.	(Algarni 2019) USA	✓	✓											✓			✓	✓	✓	✓							✓	✓	✓									
2.	(Broadhurst, Skinner, Sifniotis, Matamoros-Macias, & Ipsen 2019) Australia	✓	✓																								✓	✓	✓	✓			✓			✓	✓	
3.	(Williams & Polage, 2019) USA	✓	✓															✓									✓	✓	✓									
4.	(Kleitman, Law, & Kay, 2018) Australia	✓	✓								✓			✓	✓	✓				✓							✓	✓										
5.	(Williams, Hinds, & Joinson, 2018) UK	✓	✓																	✓																		
6.	(Vishwanath et al., 2018) USA																																					
7.	(Butavicius et al. 2017) Australia	✓	✓																								✓	✓	✓									
8.	(Butavicius, Parsons, Pattinson, & McCormac, 2015) Australia	✓	✓																	✓							✓	✓	✓									
9.	(Bojmaeh, 2015) UK	✓	✓	✓		✓			✓																													
10.	(Buckley, Nurse, Legg, Goldsmith, & Sadie, 2014)		✓		✓		✓																															
11.	(Luo et al., 2013) USA	✓	✓																✓	✓		✓				✓												
12.	(Arachchilage & Love, 2013) UK	✓	✓	✓				✓	✓	✓	✓	✓	✓														✓	✓				✓						
13.	(Wang et al., 2012) USA	✓	✓											✓	✓		✓	✓		✓	✓						✓	✓										
14.	(Vishwanath et al., 2011) USA	✓	✓						✓					✓				✓		✓	✓					✓								✓				
15.	(Sheng et al., 2010) USA	✓	✓				✓							✓													✓	✓	✓		✓			✓				✓
16.	(Liang & Xue, 2010) USA	✓	✓	✓				✓	✓	✓	✓	✓	✓														✓	✓				✓						
17.	(Liang & Xue, 2009) USA	✓	✓	✓				✓	✓	✓	✓	✓	✓														✓	✓				✓						
18.	(Huber et al., 2009) Sweden		✓			✓														✓							✓	✓										
19.	(Bakhshi, Papadaki, & Furnell, 2009) UK		✓																																			
20.	(Kumaraguru et al., 2009) USA	✓	✓				✓																				✓	✓		✓								
21.	(Kumaraguru et al., 2008) Portugal	✓	✓				✓																				✓											

	Constructs	Outcome Behaviour		Behavioural Intention	Organizational Factors			Coping Appraisal				Threat Appraisal		Knowledge			Elaboration		Attack Factors		Motivated to Process			Ability to Process			Demographic Factors											
	List of Relevant Empirical/Theoretical Studies: (Authors, Date) Country	Secure	Insecure	Threat Avoidance	Policies	Technology Controls	Training & Awareness	Response Efficacy	Self-Efficacy	Response Cost	Threat Detection	Perceived Vulnerability	Perceived Severity	Threat Domain	Detection Cues	Determinants of Trust	Central Route	Peripheral Route	Quality of Argument	Persuasive Cues	Involvement	Need For Cognition	Responsible	Distraction	Emotions	Pressure	Gender	Age	Education Level	Role	Years on Internet	Hours on Internet	Computer Skill	E-mail Load	Email Responsiveness	Online Services Usage	Prior Victimization	Risk Propensity
22.	(Egelman et al., 2008) USA		✓			✓																					✓	✓								✓		
23.	(Workman et al., 2008) USA		✓					✓	✓	✓		✓	✓																									
24.	(Workman, 2008a) USA		✓									✓	✓							✓					✓													
25.	(Workman, 2008b) USA		✓																	✓					✓		✓	✓	✓								✓	
26.	(Workman, 2007) USA		✓									✓	✓							✓					✓		✓	✓	✓									
27.	(Sheng et al., 2007) USA	✓	✓				✓																				✓	✓	✓		✓	✓				✓		
28.	(Kumaraguru, Rhee, Sheng, et al., 2007) USA	✓	✓				✓																															
29.	(Kumaraguru, Rhee, Acquisti, et al., 2007) USA	✓	✓				✓																															
30.	(Kumaraguru, Sheng, et al., 2007) USA	✓	✓				✓																				✓	✓	✓		✓	✓						
31.	(Garera et al., 2007) USA		✓												✓																							
32.	(Downs et al., 2007)		✓			✓							✓	✓	✓	✓																						
33.	(Jagatic et al., 2007) USA		✓																								✓	✓										
34.	(Jakobsson et al., 2007) USA	✓	✓													✓																						
35.	(Tsow & Jakobsson, 2007) USA	✓	✓												✓	✓																						
36.	(Dodge et al., 2007) USA	✓	✓																										✓									
37.	(Downs et al., 2006) USA	✓	✓										✓		✓	✓																				✓	✓	✓
38.	(Jakobsson & Ratkiewicz, 2006) USA	✓	✓												✓	✓																						
39.	(Dhamija et al., 2006) USA	✓	✓											✓	✓	✓											✓	✓	✓			✓				✓		
40.	(Karakasiliotis et al., 2006) UK	✓	✓												✓	✓				✓																		
41.	(Jakobsson, 2005) USA		✓																	✓																		
42.	(Aytes & Connolly, 2004) USA	✓	✓				✓		✓			✓	✓	✓																								
43.	(Grazioli, 2004)	✓	✓													✓				✓																		
44.	(Friedman et al., 2002)	✓														✓																						
45.	(Fogg et al., 2001)	✓														✓											✓	✓	✓		✓					✓		

Table 3: Summary of Major Contributions from Previous Studies

	Studies	***Description***	***Major Contributions Incorporated to Present Study***
1.	Algarni (2019)	• Lab experiment • 267 participants rated a total of 4272 messages • Theory: Elaboration Likelihood Model (ELM)	• Investigated the impact of 9 different message characteristics (Correct spelling, Correct grammar, Message length, supporting picture, Supporting video, Number of likes, Number of comments, Expressive emoji, Well-organized) on users' susceptibility to social engineering • Examined susceptibility to 5 different social engineering techniques on Facebook • Correct spelling, grammar, message length, supportive picture, supportive video, number of likes and well organization positively affect susceptibility to social engineering victimization • The more security knowledge the participants had, the less susceptible they were • Gender had a significant effect on susceptibility with women being more susceptible than men • The less the time had lapsed since a user joined Facebook, the more susceptible they were • Model R^2 = 0.49
2.	Broadhurst, Skinner, Sifniotis, Matamoros-Macias, & Ipsen (2019)	• Field experiment • Over 9 months • 138 participants who were students at an Australian university	• Examined 3 types of phishing emails: generic, tailored, and spear-phishing • The study had a small sample for analysis thereby reducing the statistical power and ability to uncover significant relationships • International students and first year students at the university were found to be more susceptible
3.	Williams & Polage (2019)	• Field study • Online questionnaire completed • 178 participants at Central Washington Uni • Theory: Cialdini's Principles	• Loss-based emails were rated more persuasive and more trustworthy than reward-based emails and subsequently more likely to respond • Emails that contained particular design cues (logo and copyright), were rated more persuasive and more trustworthy than those that did not and subsequently more likely to respond
4.	Yasin et al. (2019)	• Literature Review	• Outlines different types of social engineering attacks and lists a category of indirect social engineering attacks • Examines a list of human factors that pre-dispose people so social engineering; predominantly from Cialdini's principles of influence and persuasion • Outlines various theories that have been used in literature to study social engineering
5.	Jones & Towse (2018)	• Literature Review	• Highlights prominent phishing case studies: target, US 2016 election campaigns, Google and Facebook • Examines 3 main methods that organizations can use when assessing phishing risk: vulnerability analysis, penetration testing and cyber-risk assessments • Gives an overview of studies that examine behavioural and psychological factors for susceptibility • Factors that determine phishing susceptibility most relate to the use of persuasive techniques in the email content, contextual factors that impact people when processing emails and individual characteristics that make some people more susceptible
6.	Kleitman et al. (2018)	• Field study • 150 undergraduates in psychology at University of Sydney • 40-item online email detection task	• Perceived maliciousness correlates strongly and significantly to phishing detection • Accurate definition of phishing correlates strongly and significantly to perceived maliciousness and phishing detection • Individuals who are able to assess padlock icons were able to detect phishing more • Gender and age had no significant effects on phishing susceptibility
7.	Williams, Hinds, & Joinson (2018)	• Field study	• Mean click rate, which reflected phishing susceptibility, was 19.44% • Presence of urgency and authority cues is related to an increase in likelihood to respond to phishing

	Studies	*Description*	*Major Contributions Incorporated to Present Study*
		• 9 phishing emails sent to all employees in an organization (approx. 62,000) over 6-week period • 6 focus groups held where 32 employees participated to qualitatively examine factors that determine susceptibility, mechanisms for managing phishing attacks and efficacy of training • Theory: Cialdini's Principles, Integrated Information Processing Model of Phishing Susceptibility (IIPM), Suspicion, Cognition, and Automaticity Model (SCAM), Protection Motivation Theory (PMT)	• When determining trust or suspicion, participants stated that they use techniques like hoovering over the hyperlink, looking for errors in sender address and spelling mistakes. • New employees were identified as most susceptible because they would be unfamiliar with senders • There were definite expectations on how legitimate emails should look like. • The IT support team were highlighted as important in identifying legitimacy of emails when employees were in doubt, making this a concern for small businesses without a credible IT support team. • Employees who receive large amounts of external email, for example those working in the call center, find it more difficult to determine legitimacy of emails. • The ability to easily report potential phishing attacks and provision of timely feedback was found to be most important mechanism for determining legitimacy of emails. The next important mechanism for identifying attacks was through speaking to peers in the organization to verify emails. • Respondents were uncertain about the use of technical features when identifying phishing attacks and the consequences of submitting personal information. • Majority of the participants did not consider current training initiatives to be effective and they were more a 'tick-box' exercise. They were overloaded with information regularly circulated from multiple mechanisms.
8.	Vishwanath et al. (2018)	• Field study • Undergraduate students in communication course • 2 studies conducted: one for phishing link and another for phishing attachment • 125 students targeted in phishing link study and 220 students targeted in phishing attachment study • Theory: Heuristic Systematic Model	• Model R^2 was 17% for suspicion in phishing link attack and 20% of suspicion in spear-phishing • Higher heuristic processing reduces suspicion of phishing email; while higher systematic processing increases suspicion of phishing email • Beliefs of cyber-risk are negatively related to heuristic processing; while positively related to systematic processing of link-based phishing attacks; and consequently, directly influence suspicion of phishing emails • Poor self-regulation relates to increased habitual use of email; which lowers suspicion of phishing attacks (particularly attachment attacks)
9.	Zimmerman, Friedman, Munshi, Richmond, & Jaros (2018)	• Field study (reviewing secondary empirical data)	• Examined 42 independent variables • Models need to consider individual factors (relating to a person's time in military service) and environmental factors (relating to crime rates, economy and job availability) • Measures of insider threat related to unsuitability discharge from service, subjects of criminal investigations, involvement in security incidences and loss of classified information
10.	Butavicius et al. (2017)	• Lab study • 12 emails examined • 121 participants who were students at South Australian university	• Examined 3 types of emails: genuine, phishing and spear-phishing • Model R^2 = 0.27 of phishing and 0.505 of spear-phishing • High levels of awareness were linked to better detection of phishing • Participants from countries with high levels of individualism were better at detecting phishing • Low levels of impulsivity, high levels of agreeableness and neuroticism were linked with better detection
11.	Butavicius et al. (2015)	• Lab study • 12 emails examined • 121 participants who were students at South Australian university • Theory: Cialdini's Principles	• Examined 3 types of emails: genuine, phishing and spear-phishing • Participants incorrectly judged 71% of spear-phishing emails and 37% of phishing emails to be safe showing they were more susceptible to spear-phishing attacks • Authority was the most successful social engineering strategy while social proof was the least

	Studies	***Description***	***Major Contributions Incorporated to Present Study***
12.	Chuenchujit (2016)	• Literature review of vast number of studies on phishing	• Presents a taxonomy for the current state of research on phishing categorizing: attack factors, behavioural factors, personality factors, mitigation techniques
13.	Ali (2015)	• Literature review	• Presents the state-of-art in phishing techniques and innovative defensive measures • Identifies phishing as the most current approach for social engineering and Unintentional Insider Threat
14.	Bojmaeh, 2015)	• Field study • 220 respondents from 5 large universities located in Tehran, Iran	• Self-efficacy, intention to practice, security practice care-behaviour and security practice technology have significant impact on security behaviour
15.	Mera (2015)	• Literature review	• Examines the use of policy, training and technology to mitigate end-user unintentional insider threats • Policy is a critical control but it has to be implemented and enforced properly • Security Education Training and Awareness (SETA) is a key way of increasing skills, knowledge and enabling security-positive culture within organizations • Enabling technologies are needed to automate defense against risky behaviour
16.	Buckley, Nurse, Legg, Goldsmith, & Sadie (2014)	• Examined 10 publicly available enterprise information security policies and 5 templates • Collected details of 60 cases of accidental insider threat • Correlated the policies to the cases	• 80% of the security policies addressed human error-based incidences • 80% of the policies did not mandate software training • Only 33% gave users guidance on how to prevent social engineering
17.	Ophoff et al. (2014)	• Literature review of 92 studies relating to Insider Threats • Clear literature selection criteria and systematic review process from top 50 ranked MIS journals	• Current insider threat literature can be classified in 6 high-level categories and 13 sub-categories • The category with the least amount of studies is on insider threat management indicating opportunities for future research in these areas
18.	Greitzer et al. (2014) and CERT (2013)	• Literature review of Unintentional Insider Threat research	• Unintentional Insider Threat phenomenon is under-researched • Outlines a taxonomy for Unintentional Insider Threat cases • Presents a Unintentional Insider Threat feature model with 35 cases • Categorizes factors that lead to Unintentional Insider Threat into 4 major categories: Organizational, Human, Psychosocial/Sociocultural and Demographic
19.	Tetri & Vuorinen, (2013)	• Literature review of social engineering research from the years 1996 to 2008 which included 40 studies	• Synthesizes a definition of social engineering that characterizes it using 3 dimensions regardless of the specific technique used (1) persuasion (2) fabrication (3) data gathering • Highlights deficiencies in empirical and theoretical grounding. Out of 40 studies only 5 have empirical findings and of these only 2 are grounded in theory • Emphasizes need for multi-dimensional approach when establishing factors that lead to social engineering that should explore factors relating to: (1) Target of Attack, (2) Attack Technique, (3) Attacker and (4) Organizational Setting
20.	Luo et al. (2013)	• Field study • Actual spear phishing attack that targeted 105 faculty and staff at a public university located in Southwest US • Theory: Heuristic-Systematic Model	• Susceptibility: 36% clicked link and 15% gave login credentials • Factors that have significant effect on phishing susceptibility: argument quality, source credibility, genre conformity, pretexting, less damage.

	Studies	***Description***	***Major Contributions Incorporated to Present Study***
21.	Arachchilage & Love (2013)	• Field study • Questionnaire survey with 151 participants from Brunel and Bedfordshire University in the UK • Theory: Technology Threat Avoidance Theory (TTAT)	• Model R^2: 36% of variance in Perceived Threat, 22% of variance in Avoidance Motivation, and 15% of variance in Avoidance Behaviour • Avoidance motivation significantly influences Avoidance Behaviour • Avoidance motivation is significantly determined by Perceived Threat • Perceived Threat is significantly determined by Perceived Severity and Susceptibility • Perceived Threat fully mediates influences of Perceived Susceptibility and Perceived Severity on Avoidance Motivation • Interaction between Perceived Severity and Susceptibility on Perceived Threat and Perceived Threat and Safeguard Effectiveness on Avoidance Motivation were significant • None of the demographic variables had significant effect on Avoidance Behaviour or Avoidance Motivation
22.	Wang et al. (2012)	• Field study • Online Questionnaire Survey with 267 good data used for analysis • Theory: Theory of Deception	• Message involvement increases cognitive processing effort • Attention to visceral triggers reduces cognitive processing effort and increases the susceptibility to phishing attacks • Knowledge of phishing increases attention to phishing deception indicators • Knowledge of phishing weakens (moderates) effect of attention to visceral triggers on susceptibility to phishing • Knowledge of phishing strengthens (moderates) effect of attention to phishing deception indicators on susceptibility to phishing • Attention to phishing deception indicators reduces susceptibility to phishing attacks • Cognitive processing effort did not significantly reduce susceptibility to phishing • Knowledge of phishing did not significantly reduce susceptibility to phishing
23.	Vishwanath et al. (2011)	• Field study • Examined 2 real phishing attacks sent to users at a USA university • Collected responses from students using online questionnaire survey • Received 325 responses	• Model was able to predict 48% of a person's likelihood to respond to phishing • Attention to email's source, grammar and spelling, urgency cues, and subject line increase phishing susceptibility • Urgency cues have significant negative relationship with elaboration • Elaboration reduced phishing susceptibility but the effect was not significant • Involvement had a significant positive relationship with attention to urgency cues, elaboration and also phishing susceptibility • Elaboration mediates the effect of involvement • Email load had a significant positive relationship with phishing susceptibility
24.	Sheng et al. (2010)	• Field Experiment • 1001 participants of Online Questionnaire Survey • Control group received no training • Each of the 4 experiment groups received a different type of training • Role play task used to examine phishing susceptibility when processing 14 email samples	• 90% of participants who click a phishing link also give information to phishing websites. • All forms of training reduced susceptibility to phishing by between 34%-44% • Women were significantly more susceptible to phishing than men. Mediation analysis attributed this to them having less technical knowledge and training • Youngsters in the 18-25-year age group were significantly more susceptible to phishing. Mediation analysis attributed this to less education, fewer years on the internet, lack of prior exposure to anti-phishing training, and less risk perception • For each standard deviation increase in the knowledge test score saw a 3.6% decrease in the phishing susceptibility • For each standard deviation increase in risk perception (more risk averse) score saw a 2.8% decrease in the phishing susceptibility

	Studies	*Description*	*Major Contributions Incorporated to Present Study*
25.	Liang & Xue (2010)	• Field study • Collected 152 online questionnaire responses from students • Theory: Technology Threat Avoidance Theory (TTAT)	• Perceived susceptibility and severity have significant positive effect on perceived threat and are fully mediated by perceived threat • Perceived threat has a significant positive effect on avoidance motivation • Safeguard effectiveness has a significant positive effect on avoidance motivation • Perceived threat and safeguard effectiveness have a significant negative effect on avoidance motivation • Safeguard cost has a significant negative effect on avoidance motivation • Self-efficacy has a significant positive effect on avoidance motivation • Avoidance motivation has a significant positive effect on avoidance behavior • Gender, Age and Internet Experience added as demographic variables had no significant effect on avoidance motivation and avoidance behaviour
26.	Kumaraguru et al. (2009)	• Field experiment over 28 days • Recruited 515 participants who were randomly assigned to either a control group (that received no training) or one-train treatment group or two-train treatment group	• 90% of the participants who clicked on phishing link did so within 8 hours • Participants in the one-train or two-train groups performed significantly better than those in the control condition • No significant difference between males and females in phishing susceptibility • 18-25 age group consistently more susceptible to phishing • Tech-savvy individuals are less susceptible to phishing
27.	Kumaraguru et al. (2008)	• Field experiment • Engaged 311 staff in Portuguese company and assigned them to either control group (no training) or one of two treatment groups (generic training or spear training)	• 88% of participants who clicked phishing links also submitted information on phishing websites • Training treatment groups performed significantly better than control group
28.	Egelman et al. (2008)	• Lab experiment • Engaged 60 participants in to examine effect of 4 types browser warnings on preventing phishing • Participants used their real addresses and financial information and were sent spear phishing email immediately after completing an online purchase	• 97% of the participants were susceptible to context-specific spear phishing • Significant relationship between seeing browser multiple times and ignoring them. • Significant relationship between trusting browser warnings and obeying them • Passive browser warnings (that do not interrupt users) are ineffective and have the same result as not receiving warnings
29.	Workman (2007, 2008a, 2008b)	• Field study • Had 612 participants • Collected self-reported survey responses but also observations of actual user behaviour • Theory: Elaboration Likelihood Model, Protection Motivation Theory, Cialdini's (2001) six principles of influence/persuasion	• Individuals who perceive lower threat severity or vulnerability are more susceptible to social engineering attacks • Individuals with higher normative, continuance or affective commitment are more susceptible to social engineering attacks • More trusting or obedient individuals are more susceptible to social engineering attacks
30.	Sheng et al. (2007), Kumaraguru, Rhee, Sheng, et al. (2007), Kumaraguru, Rhee, Acquisti, et al. (2007),	• Lab experiments • Non-expert study participants • Those assigned to control group did not receive training • One treatment group tested embedded training	• All training conditions significantly reduce susceptibility to phishing • The embedded training and game conditions showed the best improvement in reducing susceptibility to phishing and confidence in decision making • Conventional anti-phishing material can be effective if users read and understand

	Studies	***Description***	***Major Contributions Incorporated to Present Study***
	Kumaraguru, Sheng, et al. (2007)	• Other treatment groups tested conventional anti-phishing material	
31.	Downs et al. (2007)	• Field study • 232 respondents filled online questionnaire survey that asked them to evaluate 5 emails and determine the action to take • a role-play exercise	• High scores on knowledge on phishing (not general computer knowledge) significantly reduced phishing susceptibility • Those who had a correct interpretation of padlock icons were less susceptible to phishing
32.	Downs et al. (2006)	• Lab study • Qualitative data analysis • 20 non-expert internet users engaged in role-play exercise to process 8 mails as Pat Jones	• Just 50% of respondents had heard of 'Phishing' and many were unable to define what it meant • The younger participants were found to engage in more online activities and also those that were more risky • Cues used to identify phishing: sender addresses (95%), lock icon (85%), presence of broken images (80%), strange URLs (55%), lack of HTTPS (35%), request for sensitive financial information (55%)
33.	(Dhamija et al., 2006)	• Lab study • 22 participants asked to examine 20 fully functional websites and determine if they were legitimate or not and to explain their reasons	• 90% of participants were fooled by well-designed phishing sites • 68% of the participants disregarded pop-up warnings • 86% of participants said they had never examined website security certificates • 31% said they had never heard of the term phishing before

2.4 The Unified Multi-Dimensional Theoretical Model

This section presents a unified multi-dimensional theoretical model that is useful for determining susceptibility to Unintentional Insider Threats. This delivers two major contributions in this research work. First, the integrated model addresses the phenomenon from multiple perspectives; that is: demographic, organizational, human and attack perspectives. Such a multi-dimensional theoretical approach is largely lacking in the existing body of work (CERT, 2013; Greitzer et al., 2014; Jones & Towse, 2018; Tetri & Vuorinen, 2013; Vishwanath et al., 2011; Wang et al., 2012). Second, the model is grounded in relevant theory to provide a guiding foundation in the study of Unintentional Insider Threats. Recent studies have identified a deficiency in theory relating to Unintentional Insider Threat research (Luo et al., 2013; Tetri & Vuorinen, 2013; Wang et al., 2012; Workman, 2007).

There is scarce empirical data and poor theoretical focus on the published work relating to Unintentional Insider Threats. CERT (2013) and Greitzer et al. (2014) have pointed out that the Unintentional Insider Threat is a phenomenon that is largely under-researched. In addition, Workman (2007), Luo et al. (2013) and Tetri & Vuorinen (2013) point out that many of the studies are not empirically supported and more importantly are poorly grounded in theory. Tetri & Vuorinen (2013) conducted a literature review of social engineering literature and pointed out that only 5 of the 40 articles reviewed were backed up by empirical evidence. Of these 5, only 2 had an underlying theoretical foundation.

The literature review shared in Chapter 2 of this research confirms that theoretical foundation is truly deficient in the existing body of knowledge relating to Unintentional Insider Threats. This research reviews 75 studies and only 21 of these (28%) are grounded in theory.

In their work, Tetri & Vuorinen (2013) make a strong case for a more comprehensive approach in the research of Unintentional Insider Threats. They argue that most studies treat the threat primarily as a problem of human weakness or error. The assumption that the human is the weakest link in the case of Unintentional Insider Threats has blinded researchers from examining other factors relating to the organizational context such as policy. Tetri & Vuorinen (2013) encourage future research to pursue a more comprehensive perspective that not only considers the insider target but also studies characteristics of the attack source, attack techniques and organizational setting.

The model presented in this chapter addresses this key gap in Unintentional Insider Threat research. It does so by addressing the phenomenon using a robust theoretical foundation that looks at the issues from multiple perspectives. The model is therefore termed as a multi-dimensional theoretical model. This novel approach is expected to advance research in the area of Unintentional Insider Threats by providing a guiding theoretical foundation for future research.

This chapter begins by presenting a justification for the theories selected for inclusion in the model. It also justifies the selection of various constructs from the existing empirical body of knowledge. The next section discusses the constructs in more depth and then presents the various hypotheses that will be later tested in the process of validating the model.

2.4.1 Justification of Theoretical Foundation and Construct Selection

It is important to state that the study of Unintentional Insider Threats is markedly different from the study of Intentional Insider Threats, particularly from a theoretical perspective (CERT, 2013; Greitzer et al., 2014; Liang & Xue, 2009; Luo et al., 2013). Theories used in the study of Intentional Insider Threats cannot be directly transferred to the study of Unintentional Insider Threats.

First, because these theories are established on the premise of intentionality. This means that insiders have been given specific instructions and the theories try to explain deliberate non-compliance behaviour which is termed as the knowing-doing gap (Workman et al., 2008). In contrast, the case of Unintentional Insider Threats involves insiders who may generally lack knowledge, skills and awareness of the threat (Dhamija et al., 2006; Downs et al., 2006, 2007; Jakobsson et al., 2007; Vishwanath et al., 2011). They may have received little or no guidance from their organization on the actions they should take or behaviour they should manifest.

Second, theories used in the study of Intentional Insider Threats primarily examine acceptance behaviours as opposed to the study of avoidance behaviours as would be expected in the case of Unintentional Insider Threats (Liang & Xue, 2009, 2010). Acceptance behaviours primarily relate to adoption of prescribed virtuous information systems and practices. Avoidance behaviours primarily relate to the rejection of malicious information systems and practices. The Cybernetic Theory (Wiener, 1948) distinguishes these behaviours from a theoretical perspective. Acceptance behaviours are characterized by negative feedback loops whose intention is to close the gap between the user's current state and the desired end state. However, avoidance behaviours are characterized by positive feedback loops whose intention is to widen the gap between the user's current state and the undesired end state.

The third argument that distinguishes the theoretical foundation for Unintentional Insider Threats from that of Intentional Insider Threats is the presence of a third-party actor who uses deception to influence the behavioural outcome. Deception is used by a malicious third-party agent who intends to trick the insider into compromising the security of their information or system. The behavioural outcome in the case of Unintentional Insider Threats is therefore dependent on the ability of the insider to detect the deception in addition to being able to counter it.

Therefore, theories central to the study of Intentional Insider Threats may not be appropriate for the study of Unintentional Insider Threats. This includes theories such as Theory of Planned Behaviour (TPB), Theory of Reasoned Action (TRA), Rational Choice Theory (RCT), General Deterrence Theory (GDT) and Theory of Cognitive Moral Development.

It is important to explore theories that touch on: persuasion, deception, threat detection and threat avoidance. Once these theories are identified, they can be integrated into a consolidated multi-dimensional model.

Persuasion and Deception Theories

The first category of theories to consider is those relating to persuasion and deception. The theory selected should provide an understanding of persuasion and deception tactics that are used by attackers to deceive insiders. The theory should also explain why insiders fall for such deception.

Previous Unintentional Insider Threat literature has identified Cialdini's Principles of Influence and Persuasion as being an appropriate theoretical basis for understanding persuasion (Rusch, 1999; Workman, 2007, 2008b). Cialdini (2001) presented six constructs: authority, scarcity, liking and similarity, reciprocation, commitment and social proof. Workman (2007, 2008a, 2008b) tested the constructs of this theory empirically using: obedience and fear to examine authority; reactance to examine scarcity; trust to examine liking; normative commitment to examine reciprocation; continuance commitment to examine commitment and affective commitment to examine social proof. Results of hypothesis testing supported each of the constructs with the exception of reactant. This shows that the six constructs proposed by Cialdini are very useful in understanding tactics that can be incorporated in attacks to make them more successful.

However, the list provided by Cialdini is not exhaustive. Other studies by Bezuidenhout, Mouton, & Venter (2010) and Peltier (2006) have identified other tactics that could be equally persuasive such as emotions, overloading, manipulative relationships and diffusion of responsibility. It is therefore important to select a theory that can accommodate various attack factors regardless of technique.

A review of recent Unintentional Insider Threat literature identifies four other theories: *Interpersonal Deception Theory* (Vishwanath et al., 2011), *Theory of Deception* (Johnson et al., 2001; Vishwanath et al., 2011), *Elaboration Likelihood Model* (Rusch, 1999; Vishwanath et al., 2011; Workman, 2007, 2008b), and *Heuristic Systematic Model* (Luo et al., 2013).

The Interpersonal Deception Theory by Buller & Burgoon (1996) presents a theory that is useful for analyzing deception when it takes place in interactive contexts; mostly face-to-face encounters. A key element presented in this theory is the ability of the sender of deceptive communication to adjust their message based on responses they get from the receiver in order to make their deception successful. This theory may be useful when studying inter-personal threat scenarios such as those employing physical proximity between the attacker and insider. However, such physical proximity is rarely used during attacks because of high risk of identification and apprehension of the attacker. It is therefore important to consider another theory.

The other is the Theory of Deception by Johnson, Grazioli, Jamal, & Zualkernan (1992); Johnson, Grazioli, Jamal, & Berryman (2001); and Grazioli (2004). It is largely similar and consistent with the Interpersonal Deception Theory but differs in three (3) key ways.

First, the Theory of Deception has found application in more disciplines and contexts than the Interpersonal Deception Theory which focuses on social psychology. Of particular interest is its application in information system studies involving deception occurring over the internet (Grazioli, 2004; Grazioli & Jarvenpaa, 2001; Vishwanath et al., 2011).

Second, the Theory of Deception unlike the Interpersonal Deception Theory can be applied to instances where there is less interaction between the attacker and the insider target. This is very useful in cases of Unintentional Insider Threat because majority of the techniques do not engage insiders in highly interactive or face-to-face communication.

Third, the Theory of Deception focuses on the target's cognitive processing in explaining why people fall for deception. This is very useful because it adds a new dimension that can be useful in determining the susceptibility to Unintentional Insider Threats. Regardless of the tactics used by the attacker, this theory presents factors that can be addressed with respect to the insider. If these factors are sufficiently addressed, then the various tactics used by attackers may not be successful.

A key weakness of the Theory of Deception is its inability to distinguish the different types of cues that need to be evaluated in order to detect deception. It approaches deception from a rational perspective but does not address subjective influences such as emotions. Unintentional Insider Threat studies (Luo et al., 2013; Vishwanath et al., 2011) have shown that there are two categories that are critical in detecting deception. If a person focuses on one category over the other, the outcome is very different. The first category is the quality of issue-relevant arguments. If a person focuses on issue-relevant arguments, they are more likely to detect deception because their decision making will be objectively focused on evaluating the truth. The second category is persuasive cues. If a person focuses on persuasive cues, they are likely to be deceived because they will be subjectively biased by elements designed by the attacker to deceive them.

In order to address this key deficiency, two deception theories have been used instead. These are the Heuristic Systematic Model (Chaiken, 1980) and the Elaboration Likelihood Model (Petty & Cacioppo, 1986). These theories provide dual-processing modes covering both the objective issue-relevant arguments and the subjective persuasive cues.

In this study, the Elaboration Likelihood Model is chosen for this research instead of the Heuristic Systematic Model because it allows for a multi-dimensional evaluation of source, message, recipient and contextual factors (Petty & Wegener, 1999). A particular way this is demonstrated, unlike other deception theories, is that it seeks to understand what would make an individual (1) motivated to process deceptive communication and (2) what would interfere with their ability to process the communication objectively.

This key distinguishing feature of the Elaboration Likelihood Model provides a theoretical basis to connect with Cialdini's (2001) six principles of influence and persuasion and other tactics proposed by Bezuidenhout, Mouton, & Venter (2010) and Peltier (2006). In addition, the Elaboration Likelihood Model has been found to be useful in many more information systems studies, particularly in recent studies by Wang et al. (2012) and Vishwanath et al. (2011) that relate to the Unintentional Insider Threat. This makes it an ideal choice over the other deception theories.

Threat Detection and Threat Avoidance Theories

The second category of theories are those that relate to threat detection and threat avoidance. It is important to understand the factors that help insiders know they are under threat. Unintentional insider threats have been known to take advantage of insider’s poor perception of threat (Algarni, 2019; Kleitman et al., 2018; Vishwanath et al., 2018), particularly due to lack of situational awareness (CERT, 2013). An understanding of threat avoidance theory informs factors that help insiders avoid targeted threats.

The theory proposed for this is the Protection Motivation Theory (PMT) by Rogers (1975, 1983). The reason this theory has been chosen is because it has been empirically tested and found appropriate by previous Unintentional Insider Threat research (Algarni, 2019; Liang & Xue, 2009, 2010; Williams et al., 2018; Workman, 2007; Workman et al., 2008).

The Protection Motivation Theory (PMT) presents two key processes (1) Threat Appraisal and (2) Coping Appraisal. The threat appraisal process is decomposed into two constructs which are: perceived severity and perceived vulnerability. Liang & Xue (2009, 2010) show that these two constructs are mediated by another construct: perceived threat. An insider can only embark on protecting themselves if they perceive they are under threat. If they are oblivious to the threat, they are exposed because their guard is down.

The coping appraisal process has evolved over time to include four constructs: response efficacy, self-efficacy, perceived cost and benefit. The model originally had only the first two constructs (Rogers, 1975, 1983). Later studies added an aspect cost-benefit analysis that has been seen to affect threat avoidance choices (Herath & Rao, 2009b; Lee & Larsen, 2009; Weinstein, 1993; Workman et al., 2008).

The Protection Motivation Theory presents two more constructs; Behavioural Intention and Outcome Behaviour. The Technology Threat Avoidance Theory (TTAT) by Liang & Xue (2009, 2010) is fundamentally built upon the Protection Motivation Theory (PMT). It includes all the constructs for the Protection Motivation Theory but also extends it using additional constructs; namely: Perceived Threat, Perceived Avoidability, Avoidance Motivation and Avoidance Behaviour. The term Avoidance Motivation is used for Behavioural Intention and Avoidance Behaviour for the Outcome Behaviour.

The Technology Threat Avoidance Theory identifies the threat appraisal process to be a threat detection process. The perceived severity and perceived vulnerability (perceived susceptibility) constructs are mediated by the perceived threat construct. In addition, the theory presents the coping appraisal process as a threat avoidance process. Three constructs: response efficacy (perceived effectiveness), self-efficacy and perceived costs are mediated by perceived avoidability.

This study proposes that these two constructs from the Technology Threat Avoidance Theory be included in the unified multi-dimensional model. This is because they bring important insights that have been empirically proven in studies by Arachchilage & Love (2013) and Liang & Xue (2010). However, these constructs are renamed to be better descriptive. The perceived threat construct is renamed to threat detection and the perceived avoidability construct is renamed to threat avoidance.

The threat detection construct is also affected by the knowledge construct as informed by existing Unintentional Insider Threat literature. Empirical studies have shown that knowledge is a critical component of threat detection (Dhamija et al., 2006; Downs et al., 2006, 2007; Grazioli, 2004; Karakasiliotis et al., 2006; Kumaraguru, Sheng, et al., 2007; Vishwanath et al., 2011; Wang et al., 2012). Knowledge has been examined from different perspectives. Knowledge on the threat domain, detection cues and determinants of trust has been shown to have an effect on threat detection.

The threat avoidance construct is affected by implementation of security measures. The Unintentional Insider Threat is an organizational challenge that should be addressed using approaches recommended in best-practice frameworks. A mixture of policy, technology and people controls is required for comprehensive mitigation (ISACA, 2012; ISO, 2013; Nieles et al., 2017).

Various empirical studies have examined the effect of people, technology and policy-based controls in the mitigation of Unintentional Insider Threats. People-based controls have been mainly through security education, training and awareness. Training has been shown to significantly reduce susceptibility to Unintentional Insider Threats (Aytes & Connolly, 2004; Kumaraguru et al., 2009, 2008; Kumaraguru, Rhee, Acquisti, et al., 2007; Kumaraguru, Rhee, Sheng, et al., 2007; Kumaraguru, Sheng, et al., 2007; Sheng et al., 2010, 2007). Work has also been done to design technology controls to protect insiders form unintentional insider threats. Many organizations have placed a premium on addressing information security using technology without investing in other controls (Luo et al., 2011; Ophoff et al., 2014). Policy has also been largely ignored as pointed out by Tetri & Vuorinen (2013).

Summary of the Justification

The arguments presented so far regarding the justification of theoretical foundation and construct selection can be summarized in Table 4.

Table 4: Justification of Theoretical Foundation and Construct Selection

Theories Considered	**Theory Selected**	**Constructs from Theory**	**Rationale for Selection**
Category 1: Persuasion and Deception • Cialdini's Principles of Influence and Persuasion • Interpersonal Deception Theory • Theory of Deception • Heuristic Systematic Model • Elaboration Likelihood Model	Elaboration Likelihood Model	• Elaboration • Argument Quality • Persuasive Cues • Motivated to Process - Involvement - Responsible • Ability to Process - Distraction - Emotions - Pressure	• Dual-processing unlike Interpersonal Deception Theory or Theory of Deception • Addresses two key dimensions: (1) attack factors and (2) human cognitive processing • Accommodates Cialdini's Principles of Influence and Persuasion by examining Persuasive Cues • Empirically tested and demonstrated to be appropriate for Unintentional Insider Threat Research by previous studies
Category 2: Threat Detection and Avoidance • Protection Motivation Theory • Technology Threat Avoidance Theory	• Protection Motivation Theory	• Threat Appraisal - Perceived Severity - Perceived Vulnerability • Coping Appraisal - Response Efficacy - Self-Efficacy - Perceived Cost - Perceived Benefit • Threat Detection (from TTAT) • Threat Avoidance (from TTAT) • Behavioural Outcome	• Provides for both threat detection and threat avoidance perspectives • Widely tested and established • Empirically tested and demonstrated to be appropriate for Unintentional Insider Threat Research by previous studies
Category 3: Empirical	• Not selected from theory but from empirical studies	• Knowledge - Threat Domain - Detection Cues - Trust Determinants • Organizational Defenses - Policies - Technology - Education, Training and Awareness	• Knowledge has been shown to support detection of Unintentional Insider Threats • Best-practice frameworks require threats to be addressed by policy, technology and people controls

2.4.2 Model Constructs and Hypothesis Generation

The multi-dimensional model presented in Figure 8 illustrates how the various constructs selected for this research are organized and related in a causal-model. The model consists of twenty-two (22) independent variables and one (1) dependent variable. The dependent variable represents the behavioural outcome under investigation which is the Unintentional Insider Threat behaviour. In addition, thirteen (13) hypotheses are outlined to describe the causal relationships between the different variables. In order to clearly investigate the relationships among the main independent and the dependent variables of this study, twelve (12) demographic variables are included in the model.

The guiding principles followed in the model's development, variable selection and hypothesis generation are drawn from the Elaboration Likelihood Model (ELM) by Petty & Cacioppo (1986), Protection Motivation Theory by Rogers (1975, 1983) and from key empirical studies summarized in Section 2.4. These guiding principles are synthesized to create a new theoretical causal model on which to ground the study of unintentional insider threats. This is a major contribution of this research.

The overall goal of this multi-dimensional theoretical causal model is to outline the factors and causal relationships that determine susceptibility to Unintentional Insider Threats. Each of the twenty-two independent variables, one dependent variable, twelve demographic variables, and thirteen sets of hypotheses (which total to twenty-five hypothesis and sub-hypothesis) are described in detail hereafter.

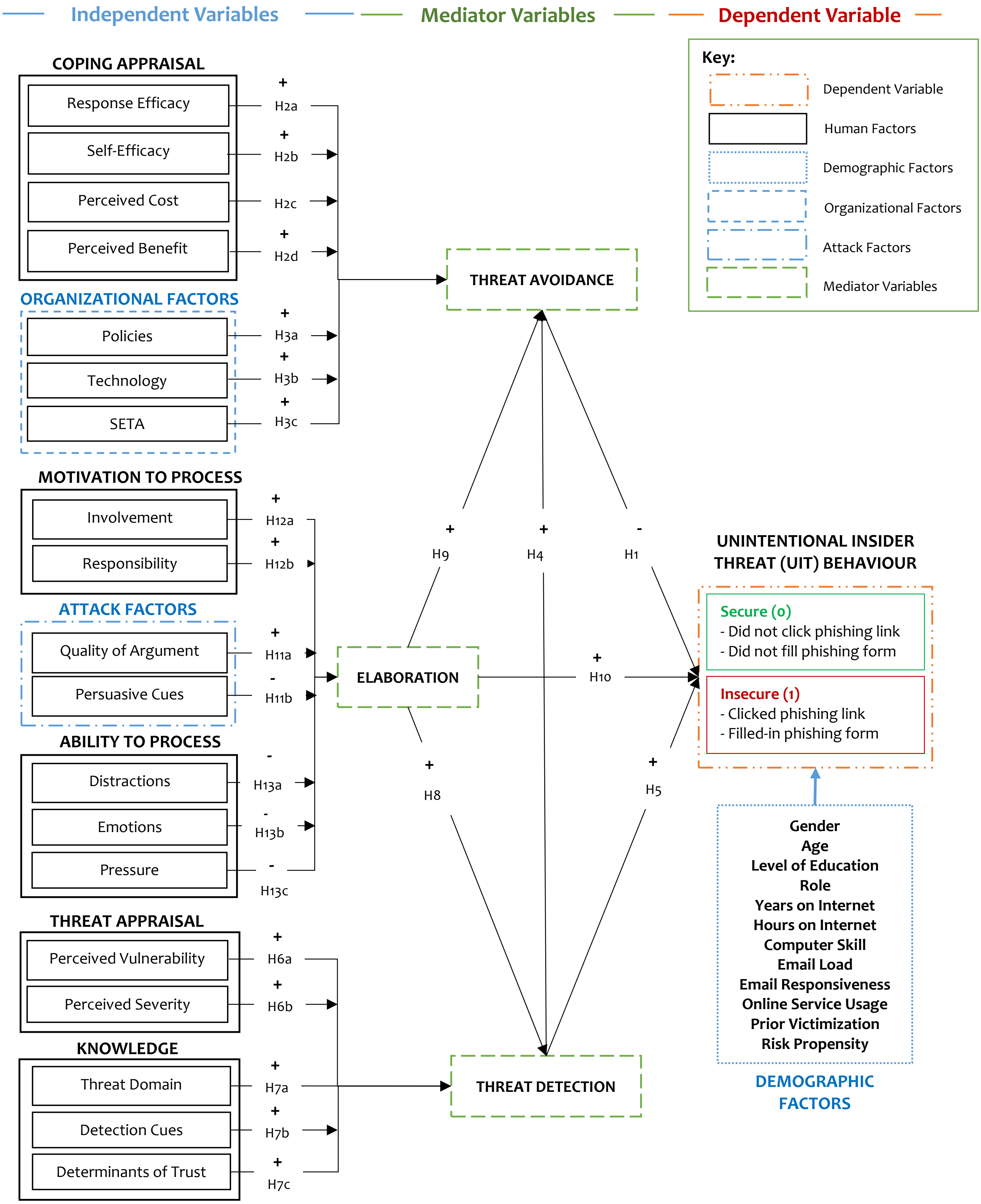


Figure 8: Proposed Multi-Dimensional Theoretical Causal Model

Outcome Variable

There is one behavioural outcome under investigation which is the Unintentional Insider Threat Behaviour outcome.

Unintentional Insider Threat Behavioural Outcome

The Unintentional Insider Threat Behaviour construct is the dependent variable in this study. Two possible outcomes are measured in this variable. The first is the insecure outcome whereby an insider takes an action that could compromise the security of their information system. The second outcome that is considered is a secure outcome whereby an insider does not take an action that could compromise the security of their information system.

The measures of the Unintentional Insider Threat Behaviour dependent variable are therefore categorical, and more specifically, dichotomous. The insecure outcome is marked with a 1 value indicating it to be 'true' as a manifestation of the unintentional insider threat while the secure outcome is marked with a 0 value indicating it to be 'false' as a manifestation of the intentional insider threat.

In this study, insiders are targeted by a deceptive social engineering attack through phishing emails. They are considered to be an Unintentional Insider Threat if they perform either one of two insecure actions determined from previous studies that address susceptibility to unintentional insider threats. These are either: (1) clicking on links on a phishing email or (2) submitting confidential information such as usernames and passwords on phishing websites (Broadhurst et al., 2019; Butavicius et al., 2017; Williams et al., 2018).

Therefore, the two instances that classify an insider as an Unintentional Insider Threat in this study are either: (1) clicking of a hyperlink in a phishing email or (2) filling in confidential details on a phishing website form. If an insider does not take any of these two actions they are not classified as an Unintentional Insider Threat.

Antecedents

The multi-dimensional model for determining susceptibility to Unintentional Insider Threats presents 22 independent variables as antecedents to the Unintentional Insider Threat behaviour outcome.

Threat Avoidance

The Threat Avoidance construct is defined as the motivation to evade Unintentional Insider threats. This construct is borrowed from the Technology Threat Avoidance Theory by Liang & Xue (2009, 2010) where they specifically examine the avoidance motivation variable. This research examines threat avoidance as a behavioural intention that is scored using self-reported measures as is presented in studies by Arachchilage & Love (2013) and Liang & Xue (2010). This separates the self-reported measures of threat avoidance from what is the directly observed Unintentional Insider Threat behaviour outcome.

Therefore, this study hypothesizes that:

Hypothesis 1: *Threat Avoidance* has a negative and significant effect on the *Unintentional Insider Threat Behaviour Outcome*.

Coping Appraisal

The Coping Appraisal construct is borrowed from the Protection Motivation Theory by Rogers (1975, 1983) and it measures a person's evaluation of: (1) response efficacy – which is the effectiveness of the recommended protective response; (2) self-efficacy – which is their ability to execute the recommended protective response; (3) perceived cost and (4) perceived benefit of recommended protective responses.

Previous studies examining protective behaviour in information security contexts (Herath & Rao, 2009b; Lee & Larsen, 2009; Liang & Xue, 2010; Workman et al., 2008) have shown that coping appraisal has a positive effect on threat avoidance. Therefore, this study hypothesizes that:

Hypothesis 2: *Coping Appraisal* has a positive and significant effect on *Threat Avoidance.*

Hypothesis 2 can be broken down further by considering the four components of Coping Appraisal as follows:

H2a: *Response Efficacy* has a positive and significant effect on *Threat Avoidance*.

H2b: *Self-Efficacy* has a positive and significant effect on *Threat Avoidance*.

H2c: *Perceived Response Cost* has a negative and significant effect on *Threat Avoidance.*

H2d: *Perceived Response Benefit* has a positive and significant effect on *Threat Avoidance.*

Organizational Factors

The organizational factors present an additional dimension in the analysis of the unintentional insider threat which is from an organizational perspective. This is an important step in providing a multi-dimensional theoretical model that is currently missing in the existing body of knowledge (Greitzer et al., 2014; Tetri & Vuorinen, 2013).

Three categories of protective measures are examined as part of this construct as prescribed by previous work (Ali, 2015; Allen, 2006; Applegate, 2009; Bojmaeh, 2015; Bulgurcu et al., 2010; CERT, 2013; Chuenchujit, 2016; Sheng et al., 2010). These are: (1) Polices put in place to address issues regarding acceptable and unacceptable use of information systems, (2) Technology Controls implemented in the information systems to prevent security incidents (3) Security Education Training and Awareness programs designed to impart knowledge and skills so that people can protect themselves from various information security threats. Therefore, this study hypothesizes that:

Hypothesis 3: *Organizational Factors* have a positive and significant effect on *Threat Avoidance.*

Hypothesis 3 can be further broken down by considering the three categories of protective measures that constitute organizational factors as follows:

H3a: *Policies* have a positive and significant effect on *Threat Avoidance.*

H3b: *Technology Controls* have a positive and significant effect on *Threat Avoidance.*

H3c: *Security Education Training and Awareness* have a positive and significant effect on *Threat Avoidance.*

Threat Detection

The Threat Detection Construct is defined as the extent to which a person is able to correctly perceive a danger. This construct has been studied in the context of Unintentional Insider Threats by Arachchilage & Love (2013) and Liang & Xue (2009, 2010) using the term Perceived Threat. The term 'Threat Detection' has been preferred over 'Perceived Threat' because it is consistent with the broader information security concept of intrusion detection even as relates to phishing detection (Bezuidenhout et al., 2010; Butavicius et al., 2015, 2017; Canfield, Fischhoff, & Davis, 2016; Levine, 2014). In addition, tools developed to counter Unintentional Insider Threats employ various techniques in order to detect possible attacks (Alsharnouby, Alaca, & Chiasson, 2015; Gupta, Arachchilage, & Psannis, 2018; Mera, 2015; Raulot, 2019) . These studies show that when a person detects a threat, they act in a way to avoid the threat. In addition, this is expected to have a direct effect on the Unintentional Insider Threat Behavioural outcome. Therefore, this study hypothesizes that:

Hypothesis 4: *Threat Detection* has a positive and significant effect on *Threat Avoidance.*

Hypothesis 5: *Threat Detection* has a negative and significant effect on the *Unintentional Insider Threat Behaviour Outcome.*

Threat Appraisal

The Threat Appraisal construct is adopted from the Protection Motivation Theory by Rogers (1975, 1983). It is defined as the extent to which a person feels at risk of harm due to an unpleasant situation (termed as a threat). This is determined by their (1) perceived severity of the threat - the level of harm that would result from the threat and (2) perceived vulnerability of the threat – the possibility that the threat could occur to them. These factors have been studied in relation to the Unintentional Insider Threat by previous studies (Arachchilage & Love, 2013; Liang & Xue, 2010; Workman, 2007; Workman et al., 2008) and have shown that the higher the perceived severity and perceived vulnerability; the more likely a person will be aware of their exposure to that threat.

Arachchilage & Love (2013) and Liang & Xue (2009, 2010) show that the effects of threat appraisal constructs are mediated by the perceived threat construct (which is renamed to threat detection in this study). Therefore, it follows that if an individual perceives a threat (threat detection) the more motivated they will be to avoid the threat (threat avoidance) by taking protective measures. Therefore, this study hypothesizes that:

Hypothesis 6: *Threat Appraisal* has a positive and significant effect on *Threat Detection*.

Hypothesis 6 can be further broken down by considering the two components of the Threat Appraisal process as follows:

H6a: *Perceived Vulnerability* has a positive and significant effect on *Threat Detection*.

H6b: *Perceived Severity* has a positive and significant effect on *Threat Detection*.

Knowledge

The Knowledge construct is defined as the level of information and skills a person acquires through experience or education that affects their understanding of a matter. This construct has been studied widely in the Unintentional Insider Threat literature from various aspects. These include knowledge relating to: terminology and threat techniques (the threat domain), cues that can be used to detect the threat (detection cues) and characteristics that can be used to distinguish legitimate communications (determinants of trust) (Dhamija et al., 2006; Downs et al., 2006; Fogg et al., 2001; Friedman et al., 2002; Garera et al., 2007; Grazioli, 2004; Jakobsson & Ratkiewicz, 2006; Jakobsson et al., 2007; Karakasiliotis et al., 2006; Sheng et al., 2010; Tsow & Jakobsson, 2007; Vishwanath et al., 2011; Wang et al., 2012).

Studies have shown that the more knowledge a person has regarding these aspects, the more likely they are able to correctly detect unintentional insider threats. Therefore, this study hypothesizes that:

Hypothesis 7: *Knowledge* has a positive and significant effect on *Threat Detection*.

Hypothesis 7 can be further broken down by considering the different kinds of knowledge as follows:

H7a: *Knowledge on Threat Domain* has a positive and significant effect on *Threat Detection*.

H7b: *Knowledge on Detection Cues* has a positive and significant effect on *Threat Detection*.

H7c: *Knowledge on Trust Determinants* has a positive and significant effect on *Threat Detection*.

Elaboration

The Elaboration Likelihood Model (ELM) by Petty & Cacioppo (1986) presents the Elaboration construct and defines it as the extent to which a person cognitively evaluates a persuasive message by paying attention to the issue-relevant arguments as opposed to paying attention to distracting persuasive cues. There are two levels of Elaboration given, namely: High Elaboration and Low Elaboration.

High Elaboration refers to instances where more cognitive effort is given to scrutinize a persuasive message based on issue-relevant arguments presented regarding an issue. This seeks to objectively sift truth from fallacy.

Low Elaboration refers to cases where less cognitive effort is dedicated in evaluating a persuasive message. Instead, the validity of a message is judged subjectively based on persuasive cues, such as its packaging and attractive look and feel. The true merits of the message are not determined by examining the content or information presented.

In the context of Unintentional Insider Threats, cognitive elaboration has an effect on whether users detect threats or not. Wang et al., (2012) shows that susceptibility to phishing emails is dependent on the cognitive effort expended in processing phishing emails. They showed that the likelihood to respond to phishing emails increases with low levels of elaboration. Conversely, with high levels of elaboration users are unlikely to fall for a phishing attempt. Similarly, Vishwanath et al. (2011) in their study show a negative relationship between Elaboration and susceptibility to phishing. Unlike these previous studies, this study examines the relationship between Elaboration and Threat Detection whereby Threat Detection exerts a mediating relationship between Elaboration and susceptibility to the Unintentional Insider Threat. Therefore, this study hypothesizes that:

Hypothesis 8: *Elaboration* has a positive and significant effect on *Threat Detection*.

Hypothesis 9: *Elaboration* has a positive and significant effect on *Threat Avoidance*.

Hypothesis 10: *Elaboration* has a negative and significant effect on the *Unintentional Insider Threat Behaviour Outcome.*

Petty & Cacioppo (1986) present four antecedents to the Elaboration construct. These are: (1) Quality of Argument, (2) Persuasive Cues, (3) Motivation to Process and (4) Ability to Process. This study classifies Quality of Argument and Persuasive Cues as Attack Factors in order to address the attacker dimension.

Attack Factors

Attack Factors is a new construct that is proposed in this study. This construct is used to examine various attack characteristics that are designed into a threat by an adversary in order to make it successful. This provides an additional dimension of study that is currently lacking that focuses on the attacker (Tetri & Vuorinen, 2013; Vishwanath et al., 2011). Attack Factors incorporates two constructs that are part of the Elaboration Likelihood Model; these are: Argument Quality and Persuasive Cues.

The Quality of Argument construct defines how well a position is justified based on available evidence or set of reasons. Using this criteria, a persuasive message is objectively judged based on its validity and merit. Luo et al. (2013) examine Quality of Argument in their study and show that users are likely to become victims if phishing messages have a high argument quality. This study examines the effect Quality of Argument has on Elaboration in line with what Petty & Cacioppo (1986) propose in their model.

Persuasive cues are described as simple cues that are placed in a message in order to subjectively influence perceptions in absence of objective argument processing. Petty & Cacioppo (1986) propose that under low elaboration, people are influenced more by persuasive cues. This is because they do not actually expend effort processing the issue-relevant arguments. Various studies have examined the effect various persuasive cues have on susceptibility to Unintentional Insider Threats (Grazioli, 2004; Huber et al., 2009; Jakobsson, 2005; Karakasiliotis et al., 2006; Luo et al., 2013; Vishwanath et al., 2011; Wang et al., 2012; Workman, 2007, 2008a, 2008b). In these studies, various persuasive cues have been enumerated; such as, spelling, grammar, layout, look and feel, security padlock icons, endorsements, logos, recipient-specific information, source, subject line and genre conformity. These cues have been found to have a significant effect in persuading people to trust deceptive messages.

In many cases, these cues are an immediate way of communicating credibility without having to scrutinize the contents of a message. Attackers therefore design their attack to defeat Elaboration. Therefore, this study hypothesizes that:

Hypothesis 11: *Attack Factors* have a negative and significant effect on *Elaboration*.

Hypothesis 11 can be further broken down by specifically considering the Attack Factors as follows:

Hypothesis 11a: *Quality of Argument* has a positive and significant effect on *Elaboration*.

Hypothesis 11b: *Persuasive Cues* have a negative and significant effect on *Elaboration*.

Motivation to Process

The construct 'Motivated to Process' is described by Petty & Cacioppo (1986) as the determination a person has to examine the content of a persuasive message. Two factors are thought to affect a person's motivation to process. These are their level of involvement in the issue presented and their level of responsibility. Involvement relates to the personal relevance or vested interest someone may have regarding the matter presented in the persuasive message. The more invested they are, the more they will be motivated to process the message. Responsibility refers to the obligation a person has to handle a matter and how accountable they are to its outcomes. The more accountable they are to a matter, the more they will be motivated to process a message regarding it.

Previous studies (Vishwanath et al., 2011; Wang et al., 2012) have studied motivation to process only by examining involvement but not responsibility. This study makes an empirical contribution to the existing body of knowledge by examining the Responsibility factor in addition to the Involvement factor.

These previous studies have shown that the higher the motivation to process, the higher the elaboration and objective scrutiny of the message. Therefore, this study hypothesizes that:

Hypothesis 12: *Motivation to Process* has a positive and significant effect on *Elaboration*.

Hypothesis 12 can be further broken down by specifically considering the factors that affect Motivation to Process as follows:

Hypothesis 12a: *Involvement* has a positive and significant effect on *Elaboration*.

Hypothesis 12b: *Responsibility* has a positive and significant effect on *Elaboration*.

Ability to Process

The 'Ability to Process' construct describes the capability an individual has to examine a persuasive message. Petty & Cacioppo (1986) describe various factors that may affect a person's ability to process. These include distraction, repetition and modality of message presentation. In their work they explain that distractions require a person to exert more effort in order to examine a message. In fact, distractions often lead to low elaboration and a reliance on persuasive cues in making judgments. Repetition is examined and shown to have one of two effects. If a message is repeated moderately, it could enhance a person's ability to process. However, if done excessively it could lead to tedium and decreased ability to process. Modality of message presentation refers to the mode of delivery; for example, audio, video or print. The various modes of communication could enhance or reduce a person's ability to process mainly due to amount of information, the rate of communication and amount of time given for elaboration.

In the study of Unintentional Insider Threats, Cialdini's (2001) six principles of influence and persuasion have been shown to have an impact on a person's ability to process social engineering threats. The six principles relate to authority, scarcity, liking and similarity, reciprocation, commitment and consistency and social proof. Studies by Karakasiliotis et al. (2006) and Workman (2007, 2008a, 2008b) have shown that these factors impair judgement and cause people to be more susceptible to social engineering attacks.

On careful examination, these six principles are seen to impair the ability to cognitively process in two ways: through emotions and pressure (Algarni, 2019; Williams et al., 2018). The emotions that often come in play during social engineering attacks are fear, guilt and trust. Pressure is created by communicating a sense of urgency and by giving rewards or issuing ultimatums for a response to be given within a stipulated period of time. Luo et al. (2013) in their study hypothesized that time pressure reduces the ability to process.

Therefore, an attacker's intention is to reduce an individual's ability to process using various techniques. Conversely if a person is able to process, they are able to demonstrate high Elaboration. Therefore, this study hypothesizes that:

Hypothesis 13: *Ability to Process* has a positive and significant effect on *Elaboration*.

Hypothesis 13 can be further broken down by specifically considering the factors that affect Ability to Process as follows:

Hypothesis 13a: *Distractions* have a negative and significant effect on *Elaboration*.

Hypothesis 13b: *Emotions* have a negative and significant effect on *Elaboration*.

Hypothesis 13c: *Pressure* has a negative and significant effect on *Elaboration*.

Demographic Factors

The multi-dimensional model for determining susceptibility to unintentional insider threats proposes 12 demographic variables that can influence the relationship between the hypothesized independent and dependent variables. These 12 demographic variables are: gender, age, level of education, role, years on the internet, hours on the internet, computer skill, email load, email responsiveness, online services usage, prior victimization and risk propensity. These variables are treated as control variables because they have been shown have an influence on the outcome variable empirically but there is no strong theoretical basis for this influence. These 12 variables are better understood as demographic characteristics more than theoretical constructs. For example, gender has been found to have an influence on susceptibility to unintentional insider threats, whereby women are more susceptible to phishing attacks. However, Sheng et al. (2010) showed through mediation analysis that this gender influence is explained by the knowledge construct because women had less technical knowledge and training than men. On the other hand, it is important to note that some studies have been unable to demonstrate a relationship between demographic factors and the unintentional insider threat phenomenon (Arachchilage & Love, 2013; Dhamija et al., 2006; Kumaraguru, Sheng, et al., 2007).

1. *Gender*

Studies by Jagatic et al. (2007) and Sheng et al. (2010) were able to show that women were more susceptible to Unintentional Insider Threats than men. However, further mediation analysis by Sheng et al. (2010) showed that these differences could be explained by differences in technical knowledge and training. Women had less technical knowledge and training than men and therefore more susceptible.

2. *Age*

Studies by Sheng et al. (2010) and Kumaraguru et al. (2009) established that participants in the 18-25 years age group were more susceptible to Unintentional Insider Threats than their older counterparts. However, this age difference was explained after mediation analysis. It was found that this age group had less training, lower level of education, fewer years of experience with the internet and were less averse to risks.

3. *Level of Education*

The study by Sheng et al. (2010) found that the level of education had the most significant effect on susceptibility to Unintentional Insider Threats than all other demographic factors. In fact, education and training were able to explain effects seen from other demographic factors such as age and gender. In this research the knowledge construct is used to theoretically explain the effect of level of education on susceptibility to Unintentional Insider Threats.

4. *Role*

The study by Kumaraguru et al. (2009) took place in a university setting and the study participants were identified as either faculty, staff or students. They found that students were significantly more likely to be susceptible to the Unintentional Insider Threat than other categories. This study intends to establish the theoretical basis for this observation. Specifically, to examine if this can be explained by the Knowledge construct.

5. *Years on the Internet*

The study by Sheng et al. (2007) showed a statistically significant positive correlation between years on the internet and the ability to correctly detect phishing attempts. This translated to a lower susceptibility to Unintentional Insider Threats. This study intends to establish which theoretical constructs explain the effect of years of internet on susceptibility to Unintentional Insider Threats.

6. *Hours on the Internet*

Fogg et al. (2001) and Arachchilage & Love (2013) captured this demographic characteristic using the term internet experience. Despite this variable being measured in various studies no statistically significant result has been established of the effect of hours on the internet on susceptibility to Unintentional Insider Threats. This study will seek to establish if this demographic characteristic has any significant influence on susceptibility to Unintentional Insider Threats.

7. *Computer Skill*

Studies by Jagatic et al. (2007), Kumaraguru et al. (2009) and Sheng et al. (2010) have shown that computer-savvy individuals are less susceptible to Unintentional Insider Threats. Some studies have measured this using a knowledge quiz and have classified people who can define terms such as phishing and cookie as computer-savvy. Other studies have examined it from a skills perspective by asking users if they have ever changed browser security settings, developed a website or helped someone else fix a computer issue.

8. *Email Load*

Vishwanath et al. (2011) was able to demonstrate that the number of emails that an individual received in a day significantly increased their likelihood to respond to phishing emails. They explained that this could be due to habituation; responding out of habit instead of a consciously reasoned action. This can be examined theoretically using the 'Ability to Process' construct extracted from the Elaboration Likelihood Model. This construct considers overload as an impairment on the ability to process.

9. *Email Responsiveness*

This is a new demographic characteristic suggested for this study. It measures the extent to which a person strives to read all messages they receive and also the extent to which they aim to respond to these messages. Review of the existing body of work has not found a study that considers this measure. However, it can be a good indicator to indicate habitual response to emails or pressures to respond that could impair a person's ability to process.

10. *Online Services Usage*

The study by Downs et al. (2006) was able to show that younger people were more significantly engaged in online activities than older people. In addition, Downs et al. (2007) found that participants who had used eBay or PayPal were less likely to click on phishing links. Many of the studies that measured participant usage of online shopping and banking did not examine if these demographic characteristics had any link to susceptibility to Unintentional Insider Threats. This study intends to establish which theoretical constructs explain the effect of online services usage on susceptibility to Unintentional Insider Threats.

11. *Prior Victimization*

Workman (2008b) was able to establish that respondents who had previously been victims of social engineering were less susceptible to staged social engineering attacks. This finding was however not statistically significant. Kumaraguru, Rhee, Acquisti, et al. (2007) pointed out that one out of ten participants in a control group stated that he did not click links on emails because he had been a victim of identity theft in the past.

The Threat Appraisal construct is most likely the theoretical explanation for the effects observed from this demographic characteristic. Victims of online deception are more likely to have higher perceived vulnerability and perceived severity measures that then increase their threat detection capability.

12. Risk Propensity

Sheng et al. (2010) was able to show that the more risk-averse participants were, the less susceptible they were to Unintentional Insider Threats. Their results showed that for every standard deviation increase in risk perception there was 2.8% less susceptibility.

This study proposes to examine this demographic characteristic through the threat appraisal and threat avoidance theoretical constructs. More risk averse participants are more likely to appraise threats highly and to avoid them.

2.5 Chapter Summary

This chapter has presented the theoretical and empirical foundation for this work. It has given a brief overview of relevant literature based on the research objectives outlined in Chapter 1. Through this process the various factors that contribute to Unintentional Insider Threats have been discussed and summarized.

In addition, this chapter has presented a major contribution of this research which is a multi-dimensional model for determining susceptibility to Unintentional Insider Threats. Previous studies have shown that the Unintentional Insider Threat phenomenon is largely under-researched and the existing body of work is mostly deficient in theoretical foundation and empirical evidence. In addition, investigation of the Unintentional Insider Threat has primarily attributed susceptibility to human weaknesses as opposed to examining multiple perspectives relating to the attack and organizational setting. The choice of theory and model constructs has been justified with arguments drawn from the existing body of work. Each of the 23 model variables has been discussed and the 13 sets of hypotheses that will be examined have been outlined. The next chapter will present a discourse regarding the research design that will be used to empirically validate this model.

CHAPTER 3: METHODOLOGY

3.1 Introduction

This chapter outlines the research philosophy, strategy and design for this study. It starts by justifying the ontological, epistemological and methodological perspectives adopted for this research. It then provides an overview of the research activities and processes relating to data collection and analysis which include the research site, measurement instrument and analysis techniques. It also presents various ethical considerations that guide this study's research activities. This research has taken up a realist ontological view, positivist and objective epistemological philosophical stance. A deductive research design that employs the use of cross-sectional data captured using a questionnaire survey is selected to allow for quantitative data to be collected. Data collection takes place through a naturalistic field study where a staged phishing attack is conducted on a university population in Nairobi, Kenya. A number of descriptive and inferential analysis techniques are applied using IBM SPSS and AMOS for Structural Equation Modeling.

3.2 Research Philosophy

Research philosophy is central to the development of knowledge. The philosophy that is employed by any research consists of assumptions about the nature of knowledge and the processes that develop it. It is important to think through research philosophy because it is the foundation for choices in the research process (Saunders, Lewis, & Thornhill, 2009).

There are two main ways of thinking about research philosophy: ontology and epistemology. Each has its distinct differences and assumptions.

3.2.1 Ontological Considerations

Ontology examines how the world operates and the nature of reality. According to Fitzgerald & Howcroft (1998), there are two main competing ontological views: realist and relativist. Table 5 highlights the key differences and assumptions of each view.

Table 5: Ontological Views (Fitzgerald & Howcroft, 1998)

Realist:	**Relativist:**
Belief that the external world is made up of pre-existing perceivable structures that exist independent of a person's cognition. There is a reality that is independent of the mind.	Belief that there are multiple-realities that exist as subjective constructions of the mind. Different people will perceive situations differently based on their subjective reality. In addition, reality is socially constructed and transmitted leading to a variation across cultures.

The realist believes the external world and phenomenon under study exist independent of an individual's experience or cognitive thought and scientific methods are able to capture this reality. This means that the existence of the world, entities in the world and laws of nature are independent of people and are not created by people. The phenomenon under study are seen as facts that are observed independent of the researcher's perception.

However, the relativist believes that there are multiple realities and truth is relative to an individual's perception of it. Diversity and not consensus is upheld. Relativism acknowledges that different people may hold different interpretations of a phenomenon due to their different cognitive perceptions. They also acknowledge that these different viewpoints are acceptable and equally legitimate. This extends to the acceptance of variation of standards, ethics and truth across cultures.

3.2.2 Epistemology Considerations

Epistemology is the philosophy of the nature, source, scope, generation and justification of knowledge. It is essentially the study of knowledge. It is concerned with determining what is considered acceptable knowledge in a given area of study and the justifications given for it.

Fitzgerald & Howcroft (1998) present four competing epistemological views and these are: positivist, interpretivist, objectivist and subjectivist. Table 6 highlights the key differences and assumptions of each.

Table 6: Epistemological Views (Fitzgerald & Howcroft, 1998)

Positivist:	**Interpretivist:**
Holds the view that the world conforms to established laws of cause-and-effect. It focuses on objectivity, measurement and repeatability. Aims to generalize observations. Only observable phenomena can provide credible data.	Holds the view that there is no universal truth. A researcher understands and interprets truth from a given frame of reference. Context is important. Details that describe reality are considered credible knowledge.

Objectivist: Holds the view that it is essential that the researcher remain detached from the research study. The observation of reality should be neutral and must not be contaminated by any biases.	**Subjectivist:** Holds the view that research findings emerge from an interaction between the researcher and their research study. The researcher's values and beliefs are central mediators to the findings.

Positivist and Objectivist views go hand-in-hand and stem from a realist ontology that professes that the external world is made up of pre-existing perceivable structures that exist independent of a person's cognition. Positivism is considered the philosophical stance of the natural scientist who seeks a hard, tangible reality whose research end-product is law-like generalizations of reality. The French philosopher, Augustine Comte, is credited as the founder of the positivist doctrine (Bhattacherjee, 2012). The positivist researcher believes that only phenomenon that can be observed and quantified lead to the production of credible knowledge. The positivist research strategy involves the use of existing theory to develop a hypothesis which will be tested and confirmed or rejected thus leading to further development of knowledge. The proposed hypothesis must be logically true or empirically validated. This process is seen as iterative whereby newly generated knowledge can be theorized and tested in future research. Creswell (2003) describes the positivist approach as being logical, empirical, reductionist, cause-and-effect oriented and deterministic based on a priori theories. In addition, the objectivist view stresses that the researcher must remain neutral and should ensure there is no interference during the research study. This protects the validity of the results from bias and from being filtered by the researcher.

The Interpretivist and Subjectivist views are interrelated and grounded on a relativist ontological view which professes the existence of multiple realities, subjective constructions of reality and cultural influences. They argue that the world is too complex and diverse to be theorized using a set of definite laws. Rich insights and experiences can be lost if such complexity is generalized. Saunders et al. (2009) point out that these views are held strongly particularly in the social sciences where research studies examine people in diverse contexts and cultures. The term social actors is often used and they are distinguished from intangible objects such as buildings and machines. The term actor depicts a metaphor where people are seen to play a part on the stage of life. Their actions are guided by the interpretation of their role in a social context. In addition, the actors also interpret other people's roles based on their own understanding.

Therefore, the interpretivist seeks to comprehend the social world within the boundaries of individual perception which is subjective.

3.2.3 Justification of the Ontological and Epistemological Position

The realist, positivist and objective views are chosen as the guiding ontological and epistemological philosophies for this study. They are deemed appropriate because unlike their counterparts, they provide a well-defined structure to guide research. They assume the presence of scientific laws and theories that provide an understanding of existing conditions but also allow for future predictions of the phenomenon regardless of the social context. Table 7 contrasts the strengths and weaknesses of the chosen philosophical position.

Table 7: Strengths and Weaknesses of Selected Research Philosophies

Chosen Philosophical View: Realist, Positivist, Objective	
Strengths:	**Weaknesses:**
• Provides well-defined structure and processes to guide research	• Considered inflexible because it assumes knowledge must be quantifiable
• Theory can be generalized to many different social contexts and knowledge gained is considered universally applicable	• Disregards context and culture which may be an inaccurate or even incomplete understanding of a phenomenon
• Discovered knowledge allows for future predictions in cause-effect deduction	• There is loss of rich non-deterministic knowledge that may prevent new understanding
• Seeks to be precise and simplified (parsimony) therefore giving focused understanding	• It may be impossible for researcher to eliminate bias and to be detached from the research
• Paves way for future research by allowing other researchers to base their research on certain scientific assumptions and re-use reliable instruments and scales	• Often, there is presence of error introduced by research methods or tools that alters results

This research seeks to minimize the disadvantageous properties of the chosen philosophical views. It is argued that the Unintentional Insider Threat phenomenon can be studied with the chosen strategy because it is not purely a social phenomenon. In addition, previous studies have provided a foundation to build on. There exist reliable scales of measurement and the data collected in this study is explained in a way that is understood in the context it is generated. Care is taken to prevent bias and error from contaminating the results of the study. Guidance is sought from lessons learnt in previous studies in order to guarantee validity and reliability of research findings.

The primary aim of this research is the discovery of scientific knowledge through scientific enquiry. Objectivity is the core of scientific enquiry. Bhattacherjee (2012) defines scientific knowledge as a generalized body of knowledge that uses laws and theories to explain a phenomenon. This targets the discovery of knowledge that explains existing conditions but that can also be used to predict future outcomes using underlying causal structures.

This research presents a multi-dimensional model that is grounded in theory in order to explain the unintentional insider threat phenomenon. The unintentional insider threat is described as a scientific phenomenon because it is observed to naturally occur in the study environment and the objective of this study is to seek to explain why it occurs (Liang & Xue, 2010; Scheeres, 2008; Tetri & Vuorinen, 2013). This study provides a description of the constructs that make up the model, the synthesis of variables and hypothesized relationships from the existing body of knowledge. The research is empirical by nature and seeks to prove or disapprove hypotheses using data measured by objective tools and scientific processes. Data is quantified using scales that are checked for validity and reliability. This process yields objective results that are expected to be independent of the researcher's values or beliefs. Care is also taken to counter bias and to ensure that the researcher does not interfere with the data collection. Analysis of the data is conducted through established statistical methods and interpretation is guided along established criteria. The study processes, data and results are documented and made available for scrutiny by other scientists and this allows for validation and even reproduction of the study. In addition, this study seeks to discover new knowledge governing the Unintentional Insider Threat phenomenon and this knowledge is generalizable regardless of the culture or society that individuals operate.

3.3 Research Design

The research design provides an overall strategy to guide different research activities, particularly those relating to data collection, measurement and analysis. Saunders et al. (2009) points out that the research design provides a blueprint to ensure that research activities deliver on the research objectives in a valid and reliable manner. It is important to provide a justification for each of the research decisions and the justifications should show that the most appropriate methods were selected.

The research onion presented by Saunders et al. (2009) in Figure 9 provides a good guide for outlining and describing the research design used in this research. Section 3.2 has discussed the outer layer covering the research philosophies. This section will discuss the research approach, strategies, choices, time horizon and selected techniques and procedures. Each layer of the research onion will be discussed hereafter and a justification of the choices made given in-lieu of the various options available.

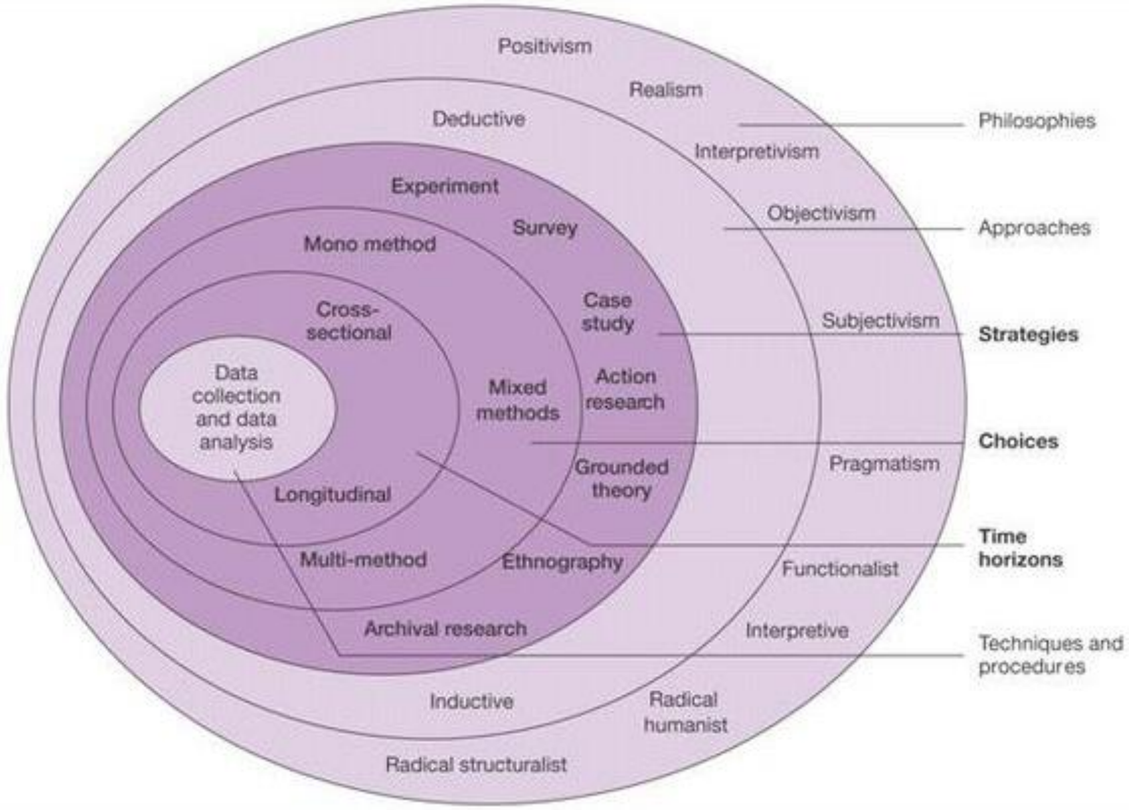


Figure 9: The Research Onion (Saunders et al., 2009)

3.3.1 Deductive Research Approach

Two distinct research approaches are identified by Saunders et al. (2009). These are deductive and inductive approaches. The deductive approach is associated with positivism and it involves use of theory, generation of hypotheses and is associated with a research strategy that collects data to test the hypotheses. Deduction is largely what is known as scientific research. On the other hand, the inductive approach is associated with the interpretivist epistemology view where a researcher first collects data then develops theory from the analysis of the data.

This research takes a positivist view and is therefore guided by the deductive research approach. This research involves 5 key steps following the deductive research approach as described by Robson (2002); (1) formulating hypotheses from theory (testable relationships between variables), (2) operationalizing the hypotheses (indicating how the variables will be measured), (3) testing the hypotheses, (4) examining the outcome (for conformity to theory or not) and (5) if necessary modification of the theory based on the findings. Any modifications to theory would need to be retested to confirm their validity.

The deductive approach seeks to explain causal relationships between variables and emphasizes scientific rigor to ensure there is no contamination of results by bias or confounding factors. In addition, the results should be generalizable and this is delivered through appropriate sampling techniques.

3.3.2 Survey Research Strategy

Seven research strategies are identified by Saunders et al. (2009) and are illustrated on the research onion process in Figure 9. These are: experiment, survey, case study, action research, grounded theory, ethnography and archival.

The experiment is considered the gold standard against which the rigor of the other research strategies is gauged. The experiment strategy has its roots in laboratory-based scientific studies that try to examine 'why' and 'how' questions regarding observed phenomena. In classical experiments, two groups are usually formed; a control group and an experiment group, and study participants are randomly assigned to either group. Care is taken to ensure there are no differences between the groups at the beginning of the study. Only the experiment group receives some form of manipulation (also called treatment) during the study to see the effect of the treatment. The control group receives none and is used as a comparison group to establish the true 'before' and 'after' effects of the treatment. The laboratory environment also allows the researchers to control their subjects so that observed changes are not generated by other confounding factors. This increases internal validity of the study outcomes. However, experiments may not work in many research scenarios especially where, for ethical reasons, subjects cannot be assigned to groups where negative results will be experienced. In addition, not many people agree to be part of experiments, therefore, the study participants may not be truly randomly selected and may represent a section of the population with unique characteristics. This lessens the external validity of experiment results.

The survey strategy often involves the administration of questionnaires to a selected sample of people. However, it can also include the use of structured interviews or observation techniques. Surveys are used to collect a large amount of data from a sizeable population in a very cost-effective way. The data is collected from each respondent in a standardized and uniform manner, making the data consistent for collective analysis. The data collected can be analyzed using both descriptive statistics (describe profile of study participants) and inferential statistics (examine relationships and suggest reasons for observations that can extend to wider population).

The case study research strategy involves empirical investigation of phenomenon within a selected real-life context using multiple types of evidence. Unlike the experiment strategy, the case study does not have clear boundaries between the phenomenon being studied and its context. The case study is often used when the researcher seeks to develop a rich understanding of the context under which the phenomenon being studied occurs. This strategy is often used for explanatory and exploratory research. Multiple sources of evidence are used and triangulation of these multiple pieces of evidence tries to confirm the results.

Action research involves a collaborative relationship between industry practitioners and researchers in the resolution of research problems that are of genuine concern to the industry. It often involves iterative processes of diagnosing the problem, planning, taking action and evaluating results. This strategy differs from the others in that it focuses on the researcher taking action in order to bring about change that will solve a research problem in a real-life context.

The grounded theory strategy places emphasis on developing and building theory. Data collection begins without any theoretical assumptions or any formulation of a theoretical framework. The collected data is analyzed for patterns that are indicative of predictions. These are tested further to see if they can be confirmed.

The ethnography strategy emanates from anthropology and is rooted in the inductive approach. Its aim is to describe the social world in which research subjects live in from their perspective. It involves the researcher spending time with the study subjects in their natural context or habitat. The data collection methods must support rich diverse data and must not be oversimplified and generalized. The researcher must be able to build a high level of trust so that the research subjects can be honest and grant full access to the researcher.

Archival research strategy is one that uses both recent and historic documents as the primary source of research data. The recorded data is generated from day-to-day activities of the research subject. The data is therefore a description of a reality that is being studied. A key to the success of archival research is an understanding of the kind of data that is needed to answer the research questions and whether that data is already available.

This research employs the survey research strategy because it allows a large amount of empirical data to be collected from study participants in a standardized manner. This data can be collected from questionnaires which have reliable scales and the research data can be used in deductive research for hypothesis testing and theory building.

3.3.3 Quantitative Data Collection Technique

With respect to the ontological and epistemological considerations and choices made, this research requires the adoption of a quantitative research approach. In addition, since the positivist view and deductive approach are chosen to guide this research, the survey data collection technique is chosen because it can provide objective, measurable and repeatable results.

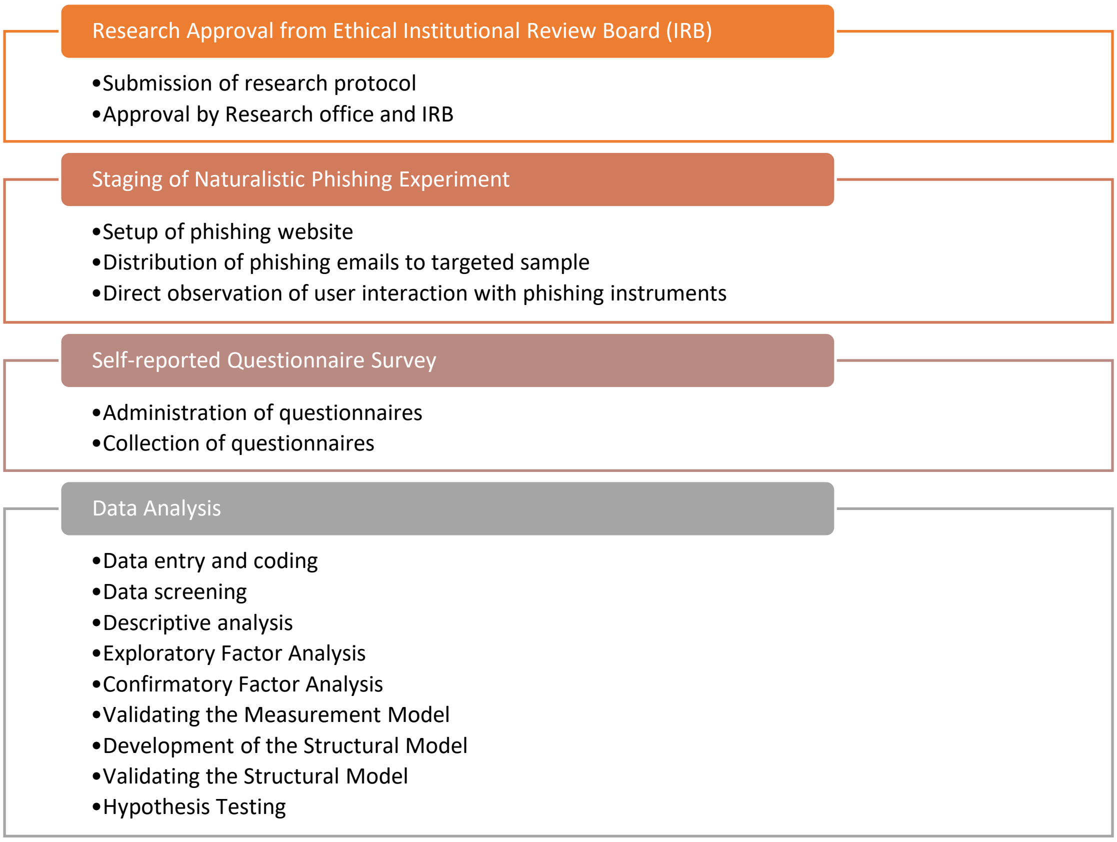
The survey research strategy is used to collect quantitative data. Quantitative methods are often employed when researchers take up the positivist view when investigating a phenomenon. The positivist view emphasizes an objective reality and the analysis of causal relationships using measured data.

3.3.4 Cross-Sectional Time-horizon

This research adopts a cross-sectional time horizon whereby the data collected represents a snapshot of the reality at a particular point in time. The study phenomena are examined at that distinct point in time. This is in line with the chosen survey research strategy that provides objective measures that can be used in deductive analysis. The longitudinal time-horizon is not chosen because it is more suited to study change and development in a cohort of study participants over a long period of time. This research does not aim to track any particular group of insiders because it does not help in the intended model and theory building.

3.3.5 Research Map

The following graphic illustrates the research map as an overview of the research journey undertaken. Figure 10 provides a visual summary of the major research methodology steps that are described hereafter.



Research Approval from Ethical Institutional Review Board (IRB)

- Submission of research protocol
- Approval by Research office and IRB

Staging of Naturalistic Phishing Experiment

- Setup of phishing website
- Distribution of phishing emails to targeted sample
- Direct observation of user interaction with phishing instruments

Self-reported Questionnaire Survey

- Administration of questionnaires
- Collection of questionnaires

Data Analysis

- Data entry and coding
- Data screening
- Descriptive analysis
- Exploratory Factor Analysis
- Confirmatory Factor Analysis
- Validating the Measurement Model
- Development of the Structural Model
- Validating the Structural Model
- Hypothesis Testing

Figure 10: Research Map

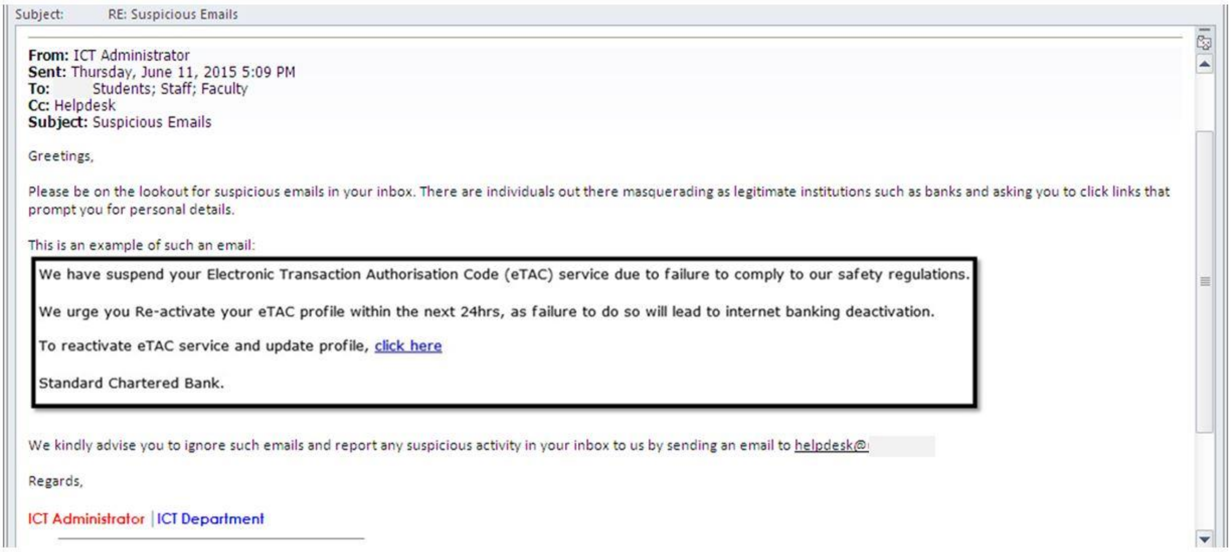
3.4 Research Setting

Decisions regarding the research setting are be guided by previous research in the field of unintentional insider threats. It is important to acknowledge that previous researchers have found it very difficult to get research approvals and co-operation to study unintentional insider threats in organizations (Bakhshi et al., 2009; Finn & Jakobsson, 2007; Huber et al., 2009; Kumaraguru et al., 2008; Vishwanath et al., 2011; Wang et al., 2012). This has been a great source of frustration that has caused some elements of research not to be conducted altogether (Huber et al., 2009) or for the research to be prematurely terminated (Bakhshi et al., 2009). In some cases organizations have refused the study data or results to be published (Kumaraguru et al., 2008).

There could be many reasons for this reluctance. Many organizations are wary of opening their doors for research due to the sensitivity of their systems and the confidential nature of their information and work practices (Burstein, 2008). They may not want their practices to be known to external parties, particularly by competitors (they might lose intellectual property or competitive edge) or even regulatory bodies (if they think their practices are deficient and may attract penalties). In addition, organizations are wary of negative publicity that may impact their bottom line, particularly due to loss of revenue or customers.

Therefore, a key criterion for selection of the research setting is obtaining a willing organization that would give approval for conducting the research, collection of sufficient data and also publication of results (Bakhshi et al., 2009). Getting a willing organization was an arduous task. Five organizations, consisting of three banks, one manufacturing company and one public utility company, were contacted over a period of 14 months to obtain approvals to conduct the research. However, approvals were not forthcoming and the loss of time was impacting on the progress of the research. The researcher therefore sought approval to conduct the research at their place of work where the approvals were given in a timely manner. The key to the approval was the element of trust bestowed on the researcher that the research would not be harmful to the organization but rather the results of the research would improve practice. There was also goodwill towards the researcher since many who were granting approvals had worked with the researcher and could vouch for their credibility. Care was taken to ensure the research was not biased by the researcher. This was done by ensuring that the research protocol and data collection instruments were independently reviewed by an Institutional Review Board (IRB) (Williams & Polage, 2019). In addition, the researcher did not score the phishing data collection instruments, rather the independent actions of the insiders would themselves objectively determine phishing susceptibility (Goel, Williams, & Dincelli, 2017).

Another key criteria for the research setting is the selection of a naturalistic environment where the unintentional insider threat phenomenon is known to occur and can be observed without alerting study subjects of the ongoing study (Vishwanath et al., 2011). The organization selected for this study had been a target of numerous social engineering attacks through phishing and they wanted assistance in addressing the issue. Many of the attacks sought to obtain the confidential data, particularly passwords, through phishing emails as illustrated in Figure 11. Other attacks sought to install malware on the information systems through malicious attachments. The organization had been hit by numerous malware infections and ransomware attacks through this social engineering vector. The organization therefore resonated with the proposed research and wanted assistance in addressing the unintentional insider threat.



Subject: RE: Suspicious Emails

From: ICT Administrator
Sent: Thursday, June 11, 2015 5:09 PM
To: [redacted] Students; Staff; Faculty
Cc: Helpdesk
Subject: Suspicious Emails

Greetings,

Please be on the lookout for suspicious emails in your inbox. There are individuals out there masquerading as legitimate institutions such as banks and asking you to click links that prompt you for personal details.

This is an example of such an email:

> We have suspend your Electronic Transaction Authorisation Code (eTAC) service due to failure to comply to our safety regulations.
>
> We urge you Re-activate your eTAC profile within the next 24hrs, as failure to do so will lead to internet banking deactivation.
>
> To reactivate eTAC service and update profile, click here
>
> Standard Chartered Bank.

We kindly advise you to ignore such emails and report any suspicious activity in your inbox to us by sending an email to helpdesk@[redacted]

Regards,

ICT Administrator | ICT Department

Figure 11: Sample Attack Received by Insiders

The organization where this research was conducted is a private university located in Nairobi, Kenya. The selection of a university as a research site in the study of unintentional insider threats has been done in several previous studies (Aldawood & Skinner, 2018; Arachchilage & Love, 2013; Broadhurst et al., 2019; Liang & Xue, 2010; Luo et al., 2013; Vishwanath et al., 2018, 2011; Wang et al., 2012). The literature review identifies 33 of the 75 studies (44%) as having been conducted at a university. Universities are a suitable research site because they are open to research and encourage the discovery of knowledge as long as the research is conducted ethically and does not harm the university systems or community (Finn & Jakobsson, 2007). In addition, universities have been the focus of recent cyberattacks with university-related scams ranking top in phishing trends (Aldawood & Skinner, 2018; Ashford, 2019a, 2019b). Universities are particularly targeted because of their vast information systems, large population of users, sometimes providing open access to members of the public and their involvement in research and innovation.

Another key criterion for the research setting is to ensure that the unintentional insider threat can be examined in a naturalistic setting. This requires staging of attacks mimicking real-life threats and targeting study subjects who are not aware of the ongoing research (Bakhshi et al., 2009; Finn & Jakobsson, 2007; Huber et al., 2009; Vishwanath et al., 2011). These staged attacks need to be conducted in a way that makes them as convincing and deceptive as would real attacks. Finn & Jakobsson (2007) explain that such naturalistic field studies need approval from the Institution Research Boards (IRB) that ensure no actual harm takes place. Such naturalistic field studies yield results with high ecological and external validity. Their results can therefore be widely generalizable beyond the context of the research site (Kumaraguru et al., 2009; Vishwanath et al., 2011; Workman, 2007, 2008a).

The research site allowed for a staged phishing attack through the research office and the university's Institutional Review Board (IRB) that granted ethical approval for the study as shown in Appendix A and B. The study was allowed to proceed without alerting the university community that they were under study. This allowed email users in the university to operate as they usually would and to interact with the staged phishing emails in a naturalistic setting. However, two senior staff in the university's information technology department were assigned to assist with the research to ensure that the phishing emails and website did not cause any actual harm to the university. The staged phishing attack did not collect any sensitive or confidential information from the study participants and neither did it transmit any malicious content.

Direct observations of the study subject's interaction with the phishing email and website was carried out. The phishing attack was staged in a way that read receipts would be received if the study subjects opened the phishing emails. In addition, a record would be kept if the study subjects clicked the hyperlink on the email. The phishing website also noted if the person filled in and submitted a form asking for the user's password. Therefore, measures were collected from the staged systems of the actual actions undertaken by the study participants without absolute reliance on self-reported data from questionnaires. Workman (2007, 2008a, 2008b) points out that direct observations are considered more objective than the subjective self-reported measures. This also builds to increase the validity of findings.

3.4.1 Population

The term population refers to all entities with the characteristics that a researcher seeks to study. It is also important to identify the unit of analysis within this population. The unit of analysis is the major entity being studied and this is commonly either an artifact, individual or group of people (Bhattacherjee, 2012).

This research is about unintentional insider threats, particularly cases of social engineering through phishing. Therefore, the targeted population is all individuals who may manifest this unintentional insider threat phenomenon. It would be difficult, if not impossible, to access the entire population of such individuals around the world. Recommendations from researchers such as Finn & Jakobsson (2007) and Jakobsson & Ratkiewicz (2006), who have done extensive work in this area, is to obtain research approvals and ethical clearance from institutions in order to stage field studies that mimic real-world attacks and to observe insider behaviour. Such staged attacks carried out as field studies have a capacity to deliver high internal and external validity which allows the findings to be generalized to wider contexts and populations.

A practical direction is therefore to select an organization willing to lend its ‘insiders’ to this study. As previously discussed in Section 4.4, getting such an organization is an arduous task. This research found it difficult to get approvals from various organizations and this had an impact on the time progress towards completion of this research. Approval was finally obtained from a private university that had experienced multiple social engineering attacks through phishing and wanted assistance in solving this issue.

Many previous studies have conducted unintentional insider threat studies at universities. In fact, 33 of the 75 studies (44%) reviewed from literature were conducted at a university. This is because universities welcome and support research as part of their core business. They not only promote research but they also have structures and resources to facilitate the research. For example, they have research and ethical review boards to review research proposals and to address any concerns that may impact the research. This ensures that the research undertaken has minimal or no risk to the population and also ensures that maximum benefit is delivered. In addition, many individuals at the institution whether students, faculty or staff are willing to participate in research activities.

The population is therefore all the insiders in a private university in Nairobi, Kenya. The unit of analysis is the insider, which is anyone who has been granted access to the organization's information systems (CERT, 2013). In the context of this study, these are all individuals in the university with active user accounts on the information systems.

3.4.2 Sampling Frame

Bhattacherjee (2012) explains that after identifying the population one has to outline the sampling frame. The sampling frame are all the accessible entities from the population from which a sample can be drawn.

In the context of this study, these are all the insiders who had active user accounts on the account management system. These are all the individuals who could be targeted by phishing attacks. The domain account management system was queried by its system administrator to provide the exact number of insiders at the time of this study. The university had a total of 8,405 insiders active on its information systems. Of these, 7,729 were students, 312 were staff members, 158 were adjunct faculty, 141 were full-time faculty, 13 were management, 9 were interns, 7 accounts were mailing list accounts and 36 could not be classified in any of these categories due to poor account metadata. Table 8 illustrates this sampling frame.

Table 8: Sampling Frame

Strata	**Number**
Students	7,729
Staff	312
Adjunct Faculty	158
Full-time Faculty	141
Management	13
Interns	9
Mailing List Users	7
Unknown	36
Total Insiders	**8,405**

3.4.3 Sampling Technique

This study employs a probability sampling technique so as to allow the results to be generalizable to the population. Bhattacherjee (2012) explains that in probability sampling, each entity in the population has a non-zero chance of being selected in the sample. In addition, random selection techniques are employed in the sampling process. This therefore ensures that sample statistics (such as mean or standard deviation) are unbiased estimates of what is in the population.

The specific technique selected is proportional stratified random sampling. The process as outlined by Bhattacherjee (2012) involves dividing the sampling frame into non-overlapping groups called strata. Thereafter a simple random sample is drawn from each strata in what is called multi-stage random sampling. In order to ensure that the strata with few members is not oversampled and the resulting sample has similar ratios for the different strata, proportional random sampling was undertaken.

3.4.4 Sample Size

The determination of sample size used the Cochran (1977) formula. It targeted a 95% confidence level and a very low margin of error at 1%. The proportion of sampling in the population was set at 50% to give maximum variability. Cochran (1977) provides a formula for calculating the sample size from large populations as the one considered in this study as follows:

$$n_o = \frac{(Z{\alpha/_2})^2 p(1-p)}{e^2}$$

Where:

n_o is the sample size

$Z{\alpha/_2}$ is the Z value at an 1-α% Confidence Interval

p is the estimated proportion of the attribute in the population

e^2 is the desired margin of error

Since the researcher was not sure about the proportion of the attribute in the population, it was set at 0.5 to give the maximum variability. The resulting sample size calculation was:

$$n_o = \frac{1.96^2 \times 0.5 \times (1-0.5)}{0.01^2}$$

$n_o = 9{,}604$

To adjust for proportions, the sample size was adjusted using the Finite Population Correction factor with the formula:

$$n = \frac{n_o N}{n_o + (N - 1)}$$

Where:

n is the actual sample size

n_o is the sample size

N is the population size

The actual sample size that was extracted from the population was therefore:

$$n = \frac{9{,}604 \times 8{,}405}{9{,}604 + (8{,}405 - 1)}$$

$$n = 4{,}483$$

Therefore, a sample size of 4,483 was extracted from the population of 8,405 insiders. To prevent under-sampling or over-sampling per strata, proportional stratified random sampling was done to determine the actual composition of the sample per strata. The numbers per strata selected for the sample are as represented in Table 9.

Table 9: Sample Size

Strata	Number	Proportion	Size in Sample
Students	7,729	91.96%	4,122
Staff	312	3.71%	166
Adjunct Faculty	158	1.88%	84
Full-time Faculty	141	1.68%	75
Management	13	0.15%	6
Interns	9	0.11%	4
Mailing List Users	7	0.08%	7
Unknown	36	0.43%	19
Total	**8,405**	**100%**	**4483**

The size in sample for each strata was then chosen using simple random sampling with the aid of a random number generator. To do this, the dataset associated with the 8,405 users were loaded onto a Microsoft Excel 2013 workbook. Each row of the workbook was associated with one user. The entries were grouped sequentially according to the strata outlined in Table 9. Next, a new column was added on the workbook to contain the random number. The random number was generated using the RAND() function entered as a formula =RAND() for every cell in the column. This ensured that each user entry was assigned a random number. After the random numbers were assigned, the entries were sorted in ascending order, while still maintaining the strata groupings. Finally, the required size in sample, say ‘n_s’, was selected by choosing the first n_s entries in each strata. These entries were transferred to a new workbook representing the selected sample dataset of 4,483 users.

3.5 Data Collection

Data in this study was collected from two sources. The first was observed behaviour from a naturalistic field study using staged social engineering attacks that mimic real threats. The second was through self-reported measures captured from questionnaire survey feedback from study participants.

3.5.1 Observations through Naturalistic Field Study

The use of a naturalistic field study incorporating staged attacks that mimic real-world attacks is the recommended method of collecting data regarding Unintentional Insider Threat behaviour (Bakhshi et al., 2009; Finn & Jakobsson, 2007; Huber et al., 2009; Vishwanath et al., 2011). This is because naturalistic studies seek to observe actual insider behaviour in its natural context. The insiders are not made aware of the ongoing study and are expected to operate the way they would normally do in the absence of the study. This protects against the Hawthorne effect (Parsons, 1974) where study subjects have been known to alter their behaviour due to the awareness that they are being studied. This behaviour modification contaminates the results of the study and compromises the validity and reliability of the study. Huber et al. (2009), Kumaraguru et al. (2009) and Workman (2007, 2008a) have also added that such naturalistic field studies have high ecological validity. Brewer & Crano (2014) explain that ecological validity is associated with studies that whose settings approximate the real-world scenarios and what is everyday life for the wider population. High ecological validity therefore enables results to be generalized to wider populations with similar real-world settings. Finn & Jakobsson (2007) have also pointed out that such naturalistic field studies are more effective than lab studies, lab experiments or IQ tests in the study of Unintentional insider threats.

Lab studies and experiments are biased due to Hawthorne effects. The study participants know they are being studied and in many cases they are primed to look out for the threat (Dhamija et al., 2006). This heightens their awareness and alters their behaviour contrary to what would have been the case in their normal day-to-day activities. In addition, the selection of participants for lab studies may also introduce bias. Most of such recruitment requires study subjects to volunteer to take part in the study. There could be unique characteristics about the type of subjects who volunteer to take part in studies and the general population. This threatens the ecological validity of the study and makes it harder to generalize the findings to real-life settings and to more diverse populations.

IQ tests have been used in the study of Unintentional insider threats to determine if insiders can identify threats from different types of material presented to them. However, Anandpara, Dingman, Jakobsson, Liu, & Roinestad (2007) point out that IQ tests are not appropriate due to numerous shortcomings. Some studies have used static screenshot images of emails and websites to see if study participants can identify which are phishing attacks. This approach is considered unsuitable because static content is devoid of many security indicators and interactive content that would be available to users in real-life settings to aid in identification of deception. Other studies have staged fully interactive websites in a lab environment which users can click links and examine content in order to make judgements. However, the participants are aware that they are being studied thereby contaminating results due to the Hawthorne effect. In addition, participants are primed to look for deception which is not what would happened in real-life settings. In fact, scoring highly in IQ tests may give participants a false sense of confidence that they are not susceptible to the threat when in actual fact they are still susceptible to well-crafted attacks that take advantage of weaknesses present in their natural context. In their study, Anandpara et al. (2007) demonstrated that IQ tests do not measure the ability to detect attacks but they only measure fear-aversion to interacting with suspicious content. In their research the participants were found to avoid interacting with content they thought to be suspicious, even if it was legitimate. The IQ tests did not reveal their capabilities and skills in detecting phishing.

Despite these advantages of using naturalistic field studies to study Unintentional insider threats, Huber et al. (2009) and Kumaraguru et al. (2009)

acknowledge that such naturalistic field studies are more difficult to conduct. It is difficult to get organizations willing to cooperate with the researcher in order to stage attacks that are as realistic, convincing and deceptive as real attacks. In addition, such studies require approvals from research and ethical review boards which may be hard to get due to associated research risks. Therefore, key to the success of such research is to identify an organization that is willing to have a naturalistic field study conducted.

3.5.2 Self-reported Measures through Questionnaire Survey

The second data collection method that was used in this research was the administration of a questionnaire survey to the study sample. The questionnaires were administered after the naturalistic field study had been completed.

The use of questionnaire surveys is very important in the study of unintentional insider threats as has been demonstrated in previous studies (Algarni, 2019; Butavicius et al., 2017; Williams & Polage, 2019). This is because many of the latent constructs and variables of interest in the study cannot be directly observed (Straub, Boudreau, & Gefen, 2004; Workman, 2007). They require the study subjects to reflect on them and to report on them particularly those relating to cognitive processes.

It is important to highlight why self-reporting alone is not an adequate data collection strategy as relates to the study of unintentional insider threats and why it has to be coupled with observed measures. Insiders may not report objectively on their behaviour. Observation allows for an objective method of identifying insiders who are susceptible to Unintentional insider threats. This objectivity increases the validity and reliability of the study and its findings.

People are known not to correctly report on their behaviour for various reasons as explored by Vishwanath et al. (2011). Firstly, they may not even be aware that they are susceptible to the threat in the first place. They may be oblivious to the fact that their actions were insecure and that they were victims of an attack. Secondly, they may also not report on their behaviour correctly because of poor memory or recollection. This is particularly true when the individual does not pay particular attention to their actions and they perform them out of habit. They may not have registered enough for them to recall well. Thirdly, the other reason for incorrect reporting is that an insider may want to cover up their behaviour because they are ashamed that they were victims.

This could be, for example, because they are in a high ranking position or status in the organization and they do not want other people to know they were not knowledgeable enough to identify and avoid an attack.

This study therefore combines the data collected from direct observations with data collected from self-reported measures so as to increase the validity and reliability of findings and also in order to provide a comprehensive understanding of the factors that contribute to the susceptibility of individuals to the Unintentional Insider Threat.

3.6 Instrument Development

This section discusses how instruments were developed to allow data collection through the two methods identified for this study. The first part examines the development of the phishing emails and phishing website that was used to in the naturalistic field study to collect data through direct observations by the researcher. The second part examines the development of a questionnaire that was used to collect self-reported data from insiders who were targeted in the naturalistic field study.

3.6.1 Phishing Instruments

The specific case of Unintentional Insider Threat that was selected for this research is social engineering through phishing. An extensive discussion justifying this selection has been provided in Section 1.1. Phishing is one attack vector that is highly prevalent and that fully demonstrates the qualities of a typical Unintentional Insider Threat. Email and websites are selected to stage the phishing attack because they are the most popularly used methods (APWG, 2017; Bakhshi et al., 2009; James, 2005; Kumaraguru, Rhee, Acquisti, et al., 2007). The use of email and websites is most popular because it has been shown to reach many targets at a very low cost. The development of the phishing emails and website was guided by recommendations and lessons learnt from previous studies by Luo et al. (2013), Arachchilage & Love (2013), Vishwanath et al. (2011) and Bakhshi et al. (2009).

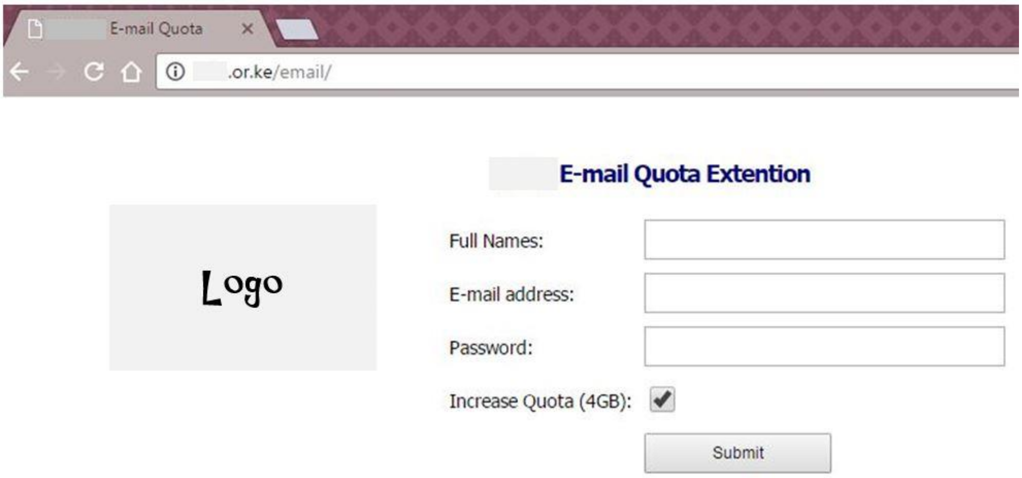
First, typical samples of phishing attacks launched against the insiders in the organization were studied. The ICT administrators who were attached to the research provided 12 samples of recent phishing attacks that had been targeted at the organization’s insiders. Characteristics that made the phishing attacks successful were identified in collaboration with the ICT administrators. The attacks that closely imitated the organization’s communication techniques and the look and feel were seen to be most deceptive. Therefore, the phishing instruments were designed to closely conform to the layout, fonts, look and feel used within the organization.

Second, a domain that imitates the organization’s domain was selected. Instead of using the registered domain ending with ac.ke, the researcher registered a domain that ended with .or.ke. The email address helpdesk@universityX.or.ke was used and the website was hosted on universityX.or.ke. This ensured that the email and website address used to conduct the attack would closely imitate the organization’s legitimate addresses but would allow for knowledgeable insiders to identify the attack by picking up an inconsistency in the addressing. This strategy is advocated by Luo et al. (2013).

The next step in the process involved the selection of a pretext scenario that would be perceived as a natural event. The pretext scenario would then guide the development of content for the phishing email and message. The guidelines by Luo et al. (2013) and Vishwanath et al. (2011) were used to guide the design of the pretext scenario. A topic that was current and relevant to the organization was selected. The organization had a limited capacity email server and consequently users were only allowed 2GB of email space. This meant that users regularly received ‘mailbox full’ notifications indicating they had exhausted their allocated quota. The pretext scenario took advantage of this and advertised an opportunity for the users to increase their allocated email quota. Time pressure was also put on the users to respond urgently in order to prevent discontinuation of service similar to the Luo et al. (2013) study.

A data collection website developed in HTML5, CSS and PHP with a MySQL database was then hosted on the registered domain and tested to ensure it ran without errors. In addition, the ICT administrators attached to the study reviewed the code and backend database to ensure that no malware was delivered, nor was any sensitive or confidential data collected and stored. This protected the insiders from actual harm as was required by directives from the research office and Institutional Review Board. Engaging the ICT administrators to review the code and backend database did not affect the realism of the study because they examined the instruments before they were deployed and not after they had collected data. The administrators did not have occasion to interact with the data collected by the phishing instruments thereafter.

Figures 12 depicts the outcome of this development process.



[redacted] E-mail Quota

[redacted].or.ke/email/

[redacted] E-mail Quota Extention

Full Names:

E-mail address:

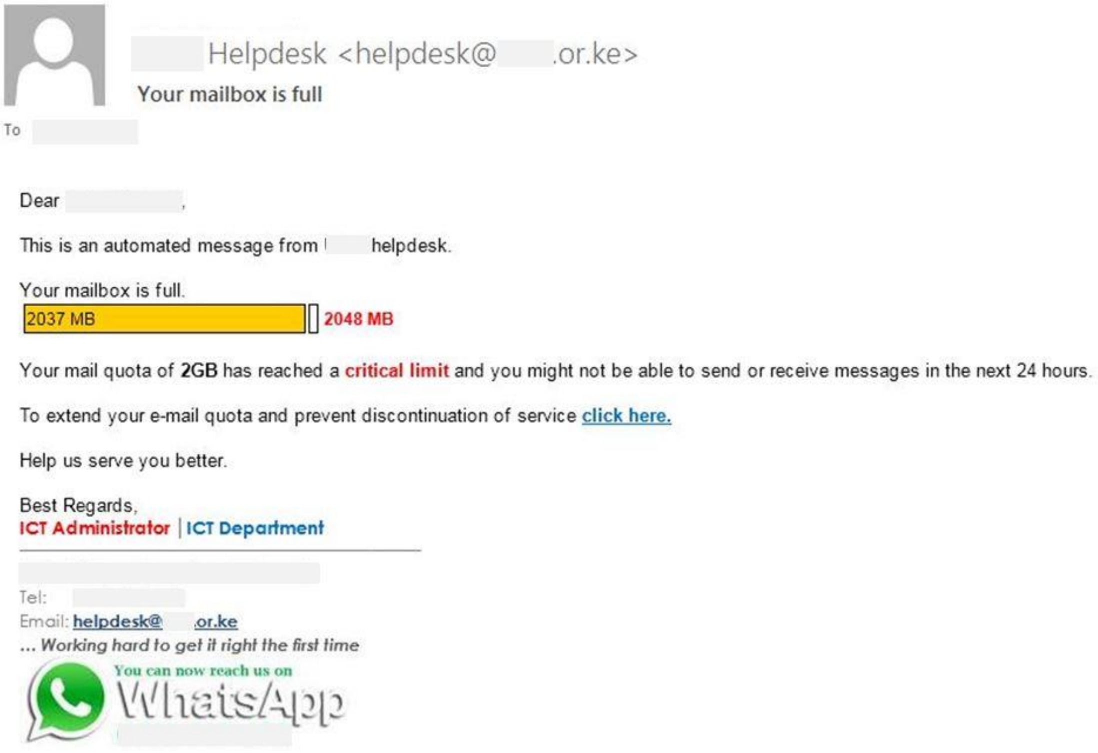
Password:

Increase Quota (4GB): ☑

Figure 12: Phishing Website

As shown in Figure 12, the phishing website was hosted on an .or.ke domain instead of the genuine ac.ke domain. In addition, the users were requested to submit in their full names, email addresses and passwords in order to get an increased email quota of 4GB. The look and feel of the webpage was designed to match that of the institution’s regular communication; including the display of the institutional logo. Any identifying information has been greyed out from the image in order to protect the identity of the institution.

Next, targeted phishing emails were sent to selected insiders. The emails were staged as spear phishing emails using the first name and surname to personalize the message. The message seemed to have been sent from the institution’s helpdesk by an ICT administrator. This imitated the means of communication commonly used by the institution when sending IT related information to the users. The email had the ‘look’ and ‘feel’ of the usual email messages from ICT administrators. It was carefully composed not to have spelling mistakes or sloppy content so that recipients do not superficially dismiss it. A mail merge template was setup using the mail merge feature on Microsoft Office Word 2013. The variable fields in the email were filled in using mail merge. These fields were: first name, last name and email address. The distribution of emails was automated using mail merge working together with Microsoft Outlook 2013. Figure 13 shows the resulting phishing email that was sent to a sample of targeted insiders.



[redacted] Helpdesk <helpdesk@[redacted].or.ke>

Your mailbox is full

To [redacted]

Dear [redacted],

This is an automated message from [redacted] helpdesk.

Your mailbox is full.

2037 MB 2048 MB

Your mail quota of **2GB** has reached a **critical limit** and you might not be able to send or receive messages in the next 24 hours.

To extend your e-mail quota and prevent discontinuation of service click here.

Help us serve you better.

Best Regards,
ICT Administrator | **ICT Department**

[redacted]

Tel: [redacted]
Email: helpdesk@[redacted].or.ke
... Working hard to get it right the first time

[redacted]

Figure 13: Phishing Email

As shown in Figure 13, the emails were staged to look like they had been sent from the institution’s helpdesk by an ICT administrator. The first name and last name of the targeted insider were used to personalize the message and stage a spear phishing attack. A sense of urgency was created by requiring the user to respond within 24 hours in order to prevent discontinuation of service. Additionally, an incentive of getting an extended email quota was put to prompt the users to take action. The users were required to act by clicking the “click here” hyperlink. The hyperlink was hidden and necessitated users to take action without providing details of the underlying web address and parameters being gathered simply by clicking the link.

These phishing instruments collected various data items for study. The phishing email had active content that tracked when the email was successfully delivered to an email address and also when the email was opened. In addition, the phishing email had a hyperlink where the words “click here” were highlighted in blue and underlined. This hyperlink did two things. First, it directed the person to the phishing website by opening their default browser and loading the phishing website’s address. Secondly, it passed on a unique identifier as a pre-filled parameter to the landing page. This means it was possible to distinctively track all the people who visited the website.

The phishing website ran active scripts that recorded to the backend database a timestamp of when the page was loaded, the identifier registered from the forwarding email and various parameters about the system accessing the page including the IP address, browser and Operating System. The source code of the background script is provided in Appendix E. This means that even if the user did not interact further with the website, just loading it gave a lot of valuable information.

The other way data was collected was when a person filled in the form on the website. This involved filling in the following details: full name, email address and password. The email address was already pre-filled if the person clicked the hyperlink from the phishing email. This communicated some level of sophistication to users that was designed to make the website more trustworthy. When a person filled in the form and clicked the submit button their password was neither captured nor transmitted as a design requirement. This prevented the capturing of confidential information and protected the institution from actual harm. The webpage also had error validation to ensure that the submit functionality did not work if the required form fields were blank.

3.6.2 Survey Instrument

The survey instrument was developed by drawing measurement items from previous studies for each construct and variable in the model. This method ensured that prior validated measures were used in the study thereby ensuring validity and reliability of measured scores (Straub et al., 2004).

3.7 Operationalization of Variables

Saunders et al. (2009) explains that the term operationalization refers to the translation of constructs used in the research model into tangible indicators that can be measured. Operationalization is central to the deductive approach because it determines how the quantitative data will be measured. The model constructs and validated measures from extant literature are presented in Table 10 and briefly explained in the following sections. In addition, the actual data collection questionnaire administered is provided in Appendix F.

3.7.1 Unintentional Insider Threat Behavioural Outcome

The Unintentional Insider Threat Behaviour is the dependent variable in this study. Two sets of measures are used. One set is made up of objective measures that capture the directly observed behaviour relating to interaction with a staged phishing email and website. The second set captures subjective self-reported measures of actions taken in relation to the email and website. This approach is similar to that used by Workman (2007) and the measures are informed by their study.

The measurement scales capture one of two possible values and is therefore a binary scale. If a study subject clicks on the hyperlink or fills in the phishing form, a 1 value is captured indicating they demonstrated the Unintentional Insider Threat Behavioural Outcome. Likewise, if they did not then a 0 value is captured.

3.7.2 Threat Avoidance

This Threat Avoidance variable is operationalized based on the studies by Liang & Xue (2009, 2010). In their studies it is measured as Avoidance Motivation but the concept is the same. It is measured using an ordinal 5-point Likert scale where values ranged from 1 is ‘not at all’ and 5 is ‘very great extent’.

3.7.3 Coping Appraisal

The Coping Appraisal construct is made up of four variables: Response Efficacy, Self-Efficacy, Perceived Cost and Perceived Benefit. Measures for these variables are informed by studies by Liang & Xue (2010), Herath & Rao, (2009b), Lee & Larsen (2009) and Workman et al. (2008). They were measured using ordinal 5-point Likert scales where values ranged from 1 is ‘strongly disagree’ and 5 is ‘strongly agree’.

3.7.4 Organizational Factors

The Organizational Factors are made up of 3 variables: Policies, Technology Controls and Security Education Training and Awareness. Measures for these variables are informed by the study by Bojmaeh (2015), Bulgurcu et al. (2010), Sheng et al. (2010) and Downs et al. (2006). They were measured using ordinal 5-point Likert scales where values ranged from 1 is ‘strongly disagree’ and 5 is ‘strongly agree’.

3.7.5 Threat Detection

The Threat Detection variable measures are informed by the studies by Arachchilage & Love, (2013) and Liang & Xue (2010). These studies used the term Perceived Threat instead but the concept is the same. They were measured using ordinal 5-point Likert scales where 1 is ‘strongly disagree’ and 5 is ‘strongly agree’.

3.7.6 Threat Appraisal

The Threat Appraisal construct is made up of two variables: Perceived Vulnerability and Perceived Severity. Measures for these variables are informed by studies by Arachchilage & Love, (2013), Liang & Xue (2010), Lee & Larsen (2009), Workman (2007) and (Downs et al. (2007). They were measured using ordinal 5-point Likert scales where Perceived Vulnerability values ranged from 1 is ‘strongly disagree’ and 5 is ‘strongly agree’ while Perceived Severity values ranged from 1 is ‘not at all’ and 5 is ‘very great extent’.

3.7.7 Knowledge

The Knowledge construct is made up of three variables that measure an individual’s knowledge on: the threat domain, detection cues and trust determinants. Measures for these variables were informed by studies by Vishwanath et al. (2011), Downs et al. (2007), Garera et al. (2007), Tsow & Jakobsson (2007), Downs et al. (2006), Dhamija et al. (2006) and Karakasiliotis et al. (2006). Threat Domain variable was measured objectively using a knowledge quiz consisting of six questions. Each question was either assigned a 1 value if the study subject got the question right or a 0 value if the study subject got the question wrong. This allowed a cumulative value of between 0 and 6 for the Threat Domain variable. The Detection Cues variable was measured using an ordinal 5-point Likert scale where values ranged from 1 is ‘strongly disagree’ and 5 is ‘strongly agree’. The Determinants of Trust variable was measured using an ordinal 5-point Likert scale where values ranged from 1 is ‘not at all’ and 5 is ‘very great extent’.

3.7.8 Elaboration

The Elaboration variable measures are informed by the studies by Wang et al. (2012), Vishwanath et al. (2011) and Petty & Cacioppo (1986). They were measured using ordinal 5-point Likert scales where 1 is ‘strongly disagree’ and 5 is ‘strongly agree’.

3.7.9 Attack Factors

The Attack Factors construct is made up of two variables: Argument Quality and Persuasive Cues. Measures for these variables are informed by studies by Luo et al. (2013), Wang et al. (2012), Vishwanath et al. (2011), Workman (2007) and Petty & Cacioppo (1986). They were measured using ordinal 5-point Likert scales where Argument Quality values ranged from 1 is 'strongly disagree' and 5 is 'strongly agree' while Persuasive Cues values ranged from 1 is 'not at all' and 5 is 'very great extent'.

3.7.10 Motivation to Process

The Motivated to Process construct is made up of two variables: Involvement and Responsibility. Measures for these variables are informed by studies by Wang et al. (2012), Vishwanath et al. (2011) and Petty & Cacioppo (1986). They were measured using ordinal 5-point Likert scales where values ranged from 1 is 'strongly disagree' and 5 is 'strongly agree'.

3.7.11 Ability to Process

The Ability to Process construct is made up of three variables: Distractions, Emotions and Pressure. Measures for these variables are informed by studies by Luo et al. (2013), Vishwanath et al. (2011), Workman (2007) and Petty & Cacioppo (1986). They were measured using ordinal 5-point Likert scales where values ranged from 1 is 'strongly disagree' and 5 is 'strongly agree'.

3.7.12 Demographic Factors

Twelve demographic variables were selected for this study based on the effects they were found to have on the dependent variable from previous studies. These control variables are: Gender, Age, Level of Education, Role, Years on the Internet, Hours on the Internet, Computer Skill, Email Load, Email Responsiveness, Online Service Usage, Prior Victimization and Risk Propensity. The measures for these variables are informed by studies by Vishwanath et al. (2011), Bulgurcu et al. (2010), Sheng et al. (2010), Kumaraguru et al. (2009), Workman (2008b), Kumaraguru, Sheng, et al. (2007), Downs et al. (2007) and Downs et al. (2006). A variety of measurement scales were used as illustrated in Table 10.

Table 10: Operationalization of Study Variables

Constructs	**Variables**	**Measurement Items**	**Possible Values**	**Scale**	**Informing Literature**
Unintentional Insider Threat	Questionnaire Self-Reported Unintentional Insider Threat Outcome Behaviour	**QSR_OB1**: Did you read this email? **QSR_OB2**: Did you click the link labelled "click here" on this email? **QSR_OB3**: Did you fill in the form presented on the website?	0: No 1: Yes	Binary	- Liang & Xue (2010) - Workman (2007) - Bakhshi et al.(2009)
	Directly Observed Unintentional Insider Threat Outcome Behaviour	**DOB_OB1**: Observed click behaviour from the website **DOB_OB2**: Observed form-fill behaviour from the website	0: No 1: Yes	Binary	- Workman (2007) - Bakhshi et al.(2009)
Threat Avoidance	Threat Avoidance	**TAV1:** My intention was to protect my computer resources **TAV2:** My intention was to protect my data	1: not at all to 5: very great extent	Ordinal 5 point Likert	- Liang & Xue (2010)
Coping Appraisal	Response Efficacy	**RE1:** Enabling security measures would protect users from similar threats **RE2:** Enabling security measures would prevent users from being deceived by similar threats **RE3:** Enabling security measures would prevent attackers from successfully launching similar threats	1: strongly disagree to 5: strongly agree	Ordinal 5 point Likert	- Liang & Xue (2010) - Herath & Rao, (2009b) - Lee & Larsen (2009) - Workman et al. (2008)
	Self-Efficacy	**SE1:** I could learn to protect myself from similar threats without much assistance **SE2:** It would be easy for me to learn security measures to protect myself from similar threats **SE3:** I can learn new computer security skills without much difficulty	1: strongly disagree to 5: strongly agree	Ordinal 5 point Likert	- Liang & Xue (2010) - Herath & Rao, (2009b) - Lee & Larsen (2009) - Workman et al. (2008)
	Response Cost	**RC1:** Taking precautions to prevent such threats would be an inconvenience **RC2:** Taking precautions to prevent such threats would be time consuming **RC3:** Taking precautions to prevent such threats would hinder my productivity	1: strongly disagree to 5: strongly agree	Ordinal 5 point Likert	- Liang & Xue (2010) - Lee & Larsen (2009) - Workman et al. (2008)
	Perceived Benefit	**PB1:** Taking precautions to prevent similar attacks would be worthwhile **PB2:** ORG-X would benefit greatly from protecting its systems from similar attacks **PB3:** Protecting myself from similar attacks would be beneficial	1: strongly disagree to 5: strongly agree	Ordinal 5 point Likert	- Workman et al. (2008)
Organizational Factors	Policies	**POL1:** I am required to know a lot about ORG-X's information security policies **POL2:** I know the regulations outlined in ORG-X's information security policies **POL3:** ORG-X's information security policies can guide me in handling such threats	1: strongly disagree to 5: strongly agree	Ordinal 5 point Likert	Bulgurcu et al. (2010)
	Technology Controls	**TC1:** ORG-X has equipped me with technology controls that can detect such threats **TC2:** ORG-X has equipped me with technology controls that can prevent such threats **TC3:** ORG-X has equipped me with technology controls that can protect me from such threats	1: strongly disagree to 5: strongly agree	Ordinal 5 point Likert	Bojmaeh (2015)
	Security Education, Training & Awareness	**SETA1:** ORG-X has made me aware of such threats **SETA2:** ORG-X has provided me with training on how to handle such threats **SETA3:** ORG-X has given me sufficient information regarding such threats	1: strongly disagree to 5: strongly agree	Ordinal 5 point Likert	- Sheng et al. (2010) - Downs et al. (2006)
Threat Detection	Threat Detection	**TD1:** I could tell this was an online attack **TD2:** I could tell someone was trying to deceive me **TD3:** I could tell that someone was trying to capture my personal details and password	1: strongly disagree to 5: strongly agree	Ordinal 5 point Likert	- Arachchilage & Love, (2013) - Liang & Xue (2010)
Threat Appraisal	Perceived Vulnerability	**PVUL1:** The chances of receiving fraudulent emails are high **PVUL2:** I am a likely target for online attacks	1: strongly disagree to	Ordinal 5 point Likert	- Arachchilage & Love, (2013)

Constructs	Variables	Measurement Items	Possible Values	Scale	Informing Literature
		PVUL3: I am likely to encounter various online attacks	5: strongly agree		- Liang & Xue (2010) - Lee & Larsen (2009) - Workman (2007)
	Perceived Severity	**Please rate how bad you think the consequences of the following actions could be on the internet** **PS1:** Opening a suspicious email **PS2:** Opening a suspicious attachment **PS3:** Clicking a suspicious hyperlink **PS4:** Loading a suspicious website **PS5:** Filling out personal details on a website **PS6:** Sharing my ORG-X username and password	1: not at all to 5: very great extent	Ordinal 5 point Likert	- Arachchilage & Love, (2013) - Liang & Xue (2010) - Lee & Larsen (2009) - Workman (2007) - (Downs et al. (2007)
Knowledge	Quiz	**Please indicate what the following words mean with regards to information security** **KQ1:** Phishing **KQ2:** Social Engineering **KQ3:** URL **KQ4:** Certificate **KQ5:** Spoofing **KQ6**: Domain *Options to select from:* A: I have never seen this word before B: I have seen this word before but I don't know what it means C: A file used to identify websites and encrypt data D: Manipulating people to compromise the security of their systems E: A name that identifies an organization's resources on the internet F: Forging the identity of a trusted entity G: Impersonation commonly through email that tricks people into sharing sensitive information H: A term for insecure websites I: Malicious Software J: A web address	**KQC:** Knowledge Quiz Count value 0 <= Integer <=6 Where KQC is sum of correct answers.	Ratio	- Vishwanath et al. (2011) - Downs et al. (2007)
	Threat Domain	**KW1:** I have sufficient knowledge regarding this type of threat **KW2:** I have sufficient knowledge regarding the consequences of this type of threat **KW3:** I have sufficient knowledge on how to detect this type of threat **KW4:** I have sufficient knowledge on how to respond to this type of threat	1: strongly disagree to 5: strongly agree	Ordinal 5 point Likert	- Vishwanath et al. (2011) - Downs et al. (2007)
	Detection Cues	**DC1:** I know how to reveal hyperlinks hidden behind text to detect such threats **DC2:** I know how to analyze web addresses to detect such threats **DC3:** I know how to analyze web certificates to detect such threats	1: strongly disagree to 5: strongly agree	Ordinal 5 point Likert	- Garera et al. (2007) - Downs et al. (2007) - Downs et al. (2006) - Dhamija et al. (2006)
	Determinants of Trust	**To what extent did you use the following characteristics or techniques to determine the trustworthiness of the email/website?** **DT1:** Consistency in logo, colors, look and feel **DT2:** Grammar and Spelling **DT3:** Personalized greeting with your names	1: not at all to 5: very great extent	Ordinal 5 point Likert	- Tsow & Jakobsson (2007) - Dhamija et al. (2006) - Karakasiliotis et al. (2006)

Constructs	Variables	Measurement Items	Possible Values	Scale	Informing Literature
		DT4: Content (e.g. reasonableness of the explanation in email and website content) **DT5:** Context (e.g. it was expected in the prevailing circumstances) **DT6:** Email address of the sender **DT7:** Contacting the ORG-X ICT helpdesk **DT8:** Asking someone (e.g. colleague, friend) **DT9:** Web address and hyperlink evaluation **DT10:** Website encryption or padlock icon **DT11:** Website certificate **DT12:** Domain registration information (e.g. from whois) **DT13:** Security tool information (e.g. anti-phishing tool integrated in email/browser)			
Elaboration	Elaboration	**ELAB1:** I made conscious effort to evaluate the email/website **ELAB2:** I took time to evaluate the email/website **ELAB3:** I carefully evaluated the email/website	1: strongly disagree to 5: strongly agree	Ordinal 5 point Likert	- Wang et al. (2012) - Vishwanath et al. (2011) - Petty & Cacioppo (1986)
Attack Factors	Argument Quality	**QA1:** I carefully scrutinized the email message before responding **QA2:** I reasoned through the explanation given in the email before responding **QA3:** I examined the reasons given in the email before responding	1: strongly disagree to 5: strongly agree	Ordinal 5 point Likert	- Luo et al. (2013) - Petty & Cacioppo (1986)
	Persuasive Cues	**Please rate to which extent the following components of the email/website influenced your response** **PC1:** Source credibility (i.e. ICT administrator) **PC2:** Personalized Greeting **PC3:** Offer to extend your mail quota **PC4:** Warning that your email service would be discontinued **PC5:** Urgency to respond within 24 hours **PC6:** Resemblance to other ORG-X emails **PC7:** Resemblance to other ORG-X websites	1: not at all to 5: very great extent	Ordinal 5 point Likert	- Luo et al. (2013) - Wang et al. (2012) - Vishwanath et al. (2011) - Workman (2007)
Motivated to Process	Involvement	**INV1:** The email seemed very relevant to me **INV2:** The email seemed very important to my work/studies **INV3:** The email seemed very applicable to my current situation	1: strongly disagree to 5: strongly agree	Ordinal 5 point Likert	- Wang et al. (2012) - Vishwanath et al. (2011) - Petty & Cacioppo (1986)
	Responsibility	**RES1:** I am answerable to communications I receive on my ORG-X email account **RES2:** I am in control of the day-to-day operation of my ORG-X email account **RES3:** I consider myself responsible for my ORG-X email account	1: strongly disagree to 5: strongly agree	Ordinal 5 point Likert	- Petty & Cacioppo (1986)
Ability to Process	Distraction	**DIST1:** There is usually a lot of activity going on around me when reading and responding to emails **DIST2:** I usually multi-task when reading and responding to emails **DIST3:** I tend to be distracted when reading and responding to emails	1: strongly disagree to 5: strongly agree	Ordinal 5 point Likert	- Petty & Cacioppo (1986)
	Emotions	**EM1:** Reading the email invoked an emotion in me (e.g. fear, anxiety) **EM2:** I responded to this email so that I would not get into trouble **EM3:** I would have felt guilty for not responding to the email	1: strongly disagree to 5: strongly agree	Ordinal 5 point Likert	- Workman (2007)
	Pressure	**PRES1:** I am usually under pressure to move on to other tasks when reading and responding to emails **PRES2:** I usually have a sense of urgency when reading and responding to emails	1: strongly disagree to 5: strongly agree	Ordinal 5 point Likert	- Luo et al. (2013) - Vishwanath et al. (2011)

Constructs	Variables	Measurement Items	Possible Values	Scale	Informing Literature
		PRES3: I tend to rush through my emails			
Demographic Factors	Gender	**GENDER:** What is your gender?	0: Male 1: Female	Binary	Sheng et al. (2010)
	Age	**AGE:** What is your age in years?	1: less than 18 years 2: 18 - 25 years 3: 26 - 35 years 4: 36 - 45 years 5: 46 - 55 years 6: above 55 years	Ordinal	Bulgurcu et al. (2010)
	Level of Education	**EDUCATION:** What is the highest level of education you have completed?	1: Primary School 2: High School 3: Diploma 4: Undergraduate Degree (Bachelor's) 5: Graduate Degree (Master's) 6: Doctoral Degree (PhD)	Ordinal	Sheng et al. (2010)
	Role	**ROLE:** What is your role at the university?	1: Student 2: Faculty/Lecturer 3: Staff 4: Other	Nominal	Kumaraguru et al. (2009)
	Year first used the internet	**YEAR_INTERNET:** Which year did you first use the internet?	1: before 1991 2: 1991-1995 3: 1996 -2000 4: 2001-2005 5: 2006-2010 6: after 2010	Ordinal	Sheng et al. (2010)
	Hours spent on the internet in a day	**HOURS_INTERNET:** How many hours do you spend on the internet in a day?	1: less than 5 2: 5-10 3: 11-15 4: 16-20 5: 21-24	Ordinal	Kumaraguru, Sheng, et al. (2007)
	Computer Skills	**COMP_SKILLS**: How would you rate your computer skills?	1: Low 2: Basic 3: Intermediate 4: Advanced 5: Expert	Ordinal	Bulgurcu et al. (2010)
	Email Load	**EL:** How many emails do you receive in your official email account in a day?	1: less than 10 2: 11-20 3: 21-30 4: 31-40	Ordinal	Vishwanath et al. (2011)

Constructs	**Variables**	**Measurement Items**	**Possible Values**	**Scale**	**Informing Literature**
			5: 41-50 6: more than 50		
	Email Responsiveness	**ER1:** I *read* all emails I receive in my ORG-X official email account **ER2:** I *respond* to all emails I need to in my ORG-X official email account	1: strongly disagree to 5: strongly agree	Ordinal	Vishwanath et al. (2011)
	Online Services	**To what extent do you use the following online services?** **OS1:** Email **OS2:** Social Media **OS3:** Online Shopping **OS4:** Online Banking	1: not at all to 5: very great extent	Ordinal 5 point Likert	- Downs et al. (2007) - Downs et al. (2006)
	Prior Victimization	**Have you ever experienced the following online threats in the past?** **PV1:** Scam **PV2:** Online Account Hijacking **PV3:** Identity Theft **PV4:** Credit/Debit Card Fraud **PV5:** Malicious software infection	0: No 1: Yes	Binary	- Workman (2008b) - Downs et al. (2006)
	Risk Propensity	**To what extent do you agree with the following statements about your risk propensity?** **RP1:** I like taking risks **RP2:** People say I am a risk taker **RP3:** I sometimes take risks that could threaten my safety	1: strongly disagree to 5: strongly agree	Ordinal 5 point Likert	- Sheng et al. (2010) - Downs et al. (2006)

3.8 Instrument Validity

Instrument validity is the assessment of the extent to which data collected by an instrument measures the intended phenomenon (Bhattacherjee, 2012; Saunders et al., 2009). It is all about ensuring that valid measures are included in the data collection instruments. Many information systems studies do not address instrument validation which can nullify findings (Straub, 1989). It is an imperative step in this research because it provides numerous benefits. The first is that it promotes research rigor that ensures high quality research deliverables. Secondly, it allows research instruments to be reused in subsequent research. The research can be extended to new contexts and heterogeneous settings allowing new knowledge to be generated in comparison to previous finings. Thirdly, validated instruments allow for the same constructs to be measured in the same way and results can be used to improve measurement techniques and in the long run the removal of confounding factors. This leads to gradual improvement of a knowledge area. Different forms of validity exist: content, face, construct, convergent and discriminant. Two will be discussed in this section and the others will be discussed in Section 3.10 on data analysis.

3.8.1 Content Validity

Content validity is concerned with the extent to which measures on an instrument are drawn from all possible measures of the phenomenon under investigation (Straub, 1989).

One key method of achieving content validity is through extensive review of extant literature on the subject and grounding of the research in theory (Saunders et al., 2009). One main undertaking of this research is to develop a multi-dimensional model grounded in theory and backed up by empirical data. The rigorous literature review presented in Chapter 2 and the model building demonstrated in Chapter 3 were undertaken in order to achieve content validity for the model's research instruments.

The other means of achieving content validity is through engaging experts who are familiar with the content universe in the research area. These experts are asked to review the instruments to ensure they provide satisfactory coverage (Straub, 1989). This research conducted a pretest of the research instruments. The pretest involved seven Information Systems professors from three different universities. They were asked to review the instruments and meetings were held with each of them to get feedback. Four of the professors were specialists in the information security domain. The other professors had experience in research and publication and were able to give valuable feedback in the improvement of the questionnaire. Additional constructs were suggested, rewording of measures and rearrangement of items were undertaken at this pretest stage. This was done until satisfactory content validity had been achieved.

3.8.2 Face Validity

Face validity is also called logical validity because it aims to ascertain that the instruments appear to be correct (Saunders et al., 2009). Critical feedback during face validity assessment is whether the research instruments make sense to the target respondents. This ensures that the respondents understand the communication as intended by the researcher and that they respond appropriately therefore capturing the correct measures. Interestingly, Bhattacherjee (2012) points out that many popular research measures lack face validity because they include very abstract constructs that may not communicate to the intended respondents.

This research conducted a pilot test of the research instruments with 32 respondents from two different universities. These universities were similar to the university where the research would be conducted. The pilot engaged students, faculty and staff at the pilot locations just like it would at the target research site. However, the pilot universities did not allow live simulation of the staged phishing attacks but only allowed static phishing instruments in the form of printed images of Figure 12 and 13 to be presented to the respondents. Feedback from the respondents enabled questions to be reworded for better clarity. In addition, feedback received on questionnaire layout necessitated for the questions to be reordered for better sequencing and flow.

To reorder the questions, a card sorting exercise was conducted which engaged two graduate students and one Information Systems professor. Cataldo, Johnson, Kellstedt, & Milbrath (1970) explain that card sorting can help refine the presentation of survey elements. The card sorting reordered the items on the questionnaire in order to provide a better flow in the ordering of the questions.

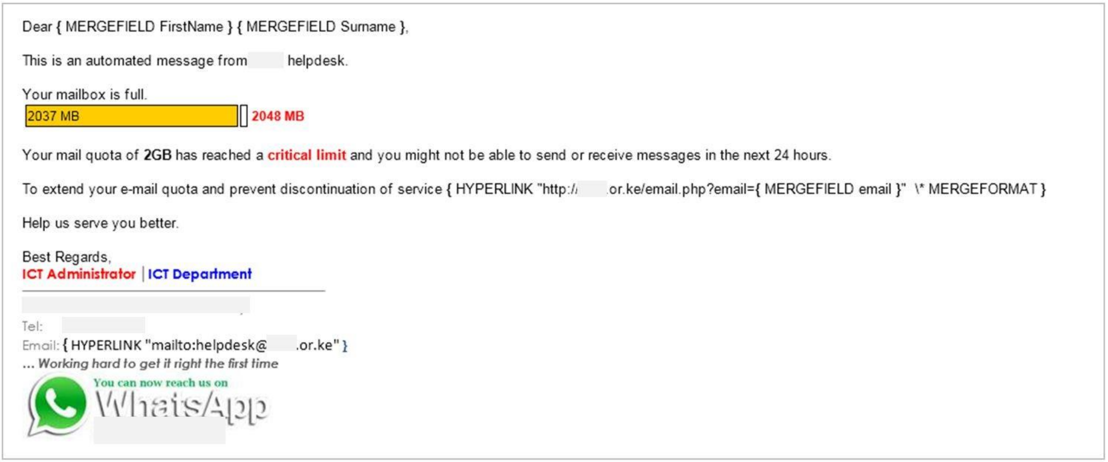
3.9 Administration of Instruments

The research instruments were only administered after they were validated as explained in Section 4.8. The instruments were pretested to ensure content validity and were also pilot tested to ensure face validity.

The first set of instruments to be administered were the phishing instruments. The phishing website was hosted on orgx.or.ke domain (which imitated orgx.ac.ke) and its interface imitated the look and feel of the organization's web content. The phishing website was designed to capture the research subject's password. If a person filled in the form and clicked the submit button their password was not captures but the backend database captured their action as an observed measure. The form had error validation to ensure that the submit functionality did not work if the form was blank.

Next, targeted phishing emails were sent to the selected population sample. The emails were staged as spear phishing emails which used the first name and surname to personalize the message. The message seemed to have been sent from the helpdesk by an ICT administrator. The phishing emails had a hyperlink leading to the phishing website. Both opening the email and clicking the phishing link were tracked as observed measure.

The administration process was automated using mail merge on Microsoft Office Word 2013 installation working with Microsoft Outlook 2013. The template is shown in Figure 14.



Dear { MERGEFIELD FirstName } { MERGEFIELD Surname },

This is an automated message from helpdesk.

Your mailbox is full.
2037 MB 2048 MB

Your mail quota of **2GB** has reached a **critical limit** and you might not be able to send or receive messages in the next 24 hours.

To extend your e-mail quota and prevent discontinuation of service { HYPERLINK "http:/ .or.ke/email.php?email={ MERGEFIELD email }" * MERGEFORMAT }

Help us serve you better.

Best Regards,
ICT Administrator | **ICT Department**

Tel:
Email: { HYPERLINK "mailto:helpdesk@ .or.ke" }
... *Working hard to get it right the first time*

Figure 14: Phishing Email Mail Merge Template

The phishing instruments were sent out for more than a month from 28th July 2016 to 6th September 2016. The phishing exercise was stopped when a prominent blogger, who was part of the sampled students, posted a comment on social media that got the university administration concerned. The ICT director had to send an alert to the entire university community informing them of the research. Thereafter the exercise was stopped. The questionnaire was then administered to all study participants who were noted to have opened the phishing email or interacted with the phishing website.

3.10 Data Entry and Coding

After the data from the naturalistic phishing exercise was obtained from the hosting servers and data from the self-reported questionnaires were received back from the study participants, they were captured on a Microsoft Excel spreadsheet and later on exported to IBM SPSS. Data was first captured on Microsoft Office Excel to allow for the data entry exercise to take place on machines other than the machine that had the licensed IBM SPSS software. In addition, the Excel data file provided some flexibility because the file could be imported onto different software for analysis as either an .xls or .csv file.

One research assistant who had been trained by the university's research office was engaged to transcribe the data from the physical questionnaires to the Microsoft Office Excel file. The code book outlined in Appendix G was used for data coding. The code book helped translate the questionnaire responses into values; for example, the value 1 was used to capture questionnaire responses indicated female. The data was then reviewed for correctness by the researcher. Next, the Microsoft Office Excel data file was imported onto IBM SPSS Statistics version 23 and a SPSS .sav file was generated. Data was coded on SPSS variable view to capture proper variable names, data types, labels, range of possible values, missing values, measurement levels and roles. The .sav file was later imported into IBM AMOS for Structural Equation Modeling analysis.

3.11 Data Analysis

Analysis of the data in this research is done using the Structural Equation Modeling (SEM) technique which is considered as a relatively new analysis technique compared to others that have been in existence for a longer time.

3.11.1 Rationale for Data Analysis Technique

There are many reasons for choosing Structural Equation Modeling for analysis. The first is that Structural Equation Modelling technique allows for the model with multiple variables and relationships to be analyzed in its entirety and not in piecemeal. Traditional analysis techniques such as correlation, regression, multiple regression and analysis of variance examine single relationships between independent and dependent variables at a time. However, it is better to test a model by examining the interplay of multiple interdependent relationships between dependent and independent variables. Structural Equation Modeling delivers such a comprehensive analysis method that analyzes the entire theory with its multiple interrelated relationships while considering all possible information (Hair, Black, Babin, & Anderson, 2009).

Another reason for choosing Structural Equation Modeling technique for analysis is that it allows for unobserved latent factors that are not measured directly to be included in the analysis. There are a number of latent factors in the multi-dimensional model that were not measured directly that need to be analyzed as outlined in Section 2.4.2. For example, coping appraisal, threat appraisal, attack quality, motivation to process and ability to process. Both observed (measured) and unobserved (latent) variables can be included in the same model for analysis.

In addition, Structural Equation Modeling is highly flexible and allows many types of relationships to be specified and analyzed. The graphical interface provided on many SEM tools allows for complex relationships between variables to be illustrated diagrammatically as illustrated in Figure 15. It is possible to illustrate path diagrams where independent variables X_1 and X_2 relate to one dependent variable Y_1 as is shown in Figure 15 (a). In addition, it is possible to modify the path diagram to illustrate an independent variable X_2 having a relationship with two different dependent variables Y_1 and Y_2 as illustrated in Figure 15 (b). Further complex relationships involving mediating variables Y_1 and Y_2 and a final dependent variable Y_3 are possible as illustrated in Figure 15 (c). All these different path diagrams are then translated by the SEM analysis engine to a set of equations that are solved simultaneously in the analysis of the model. This graphical tool support and flexibility in analysis was very desirable in this research.

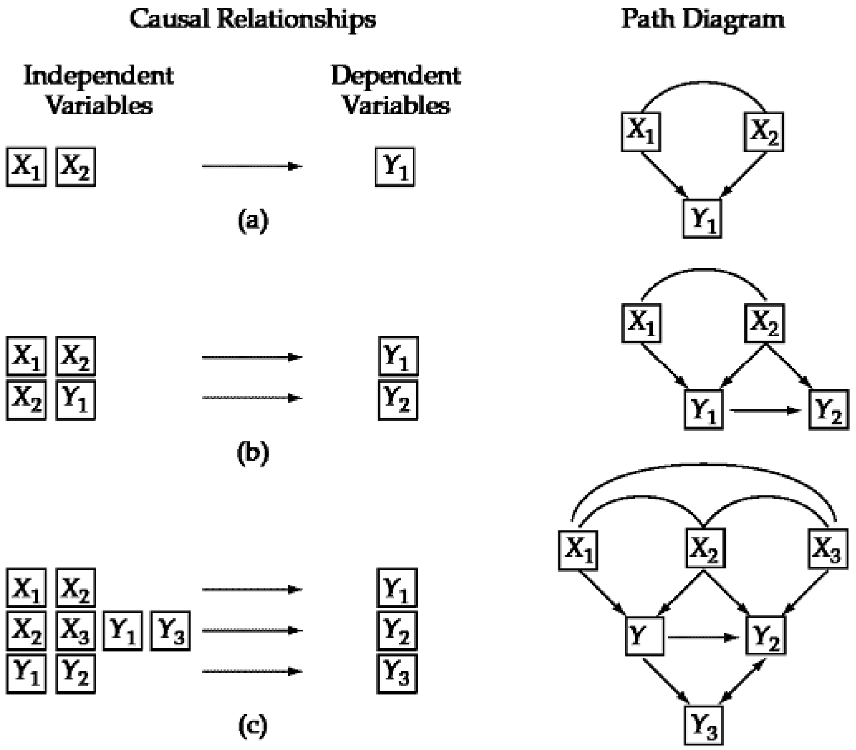


Figure 15: Complex Relationships Depicted as Path Diagrams (Hair et al., 2009)

3.11.2 Structural Equation Modeling Process Steps

Structural Equation Modeling can be considered as a combination of two distinct processes: factor analysis and path analysis. The factor analysis section is associated with the measurement model while the path analysis section is associated with the structural model. However, Hair et al. (2009) break these down to a six stage process that is illustrated in Figure 16.

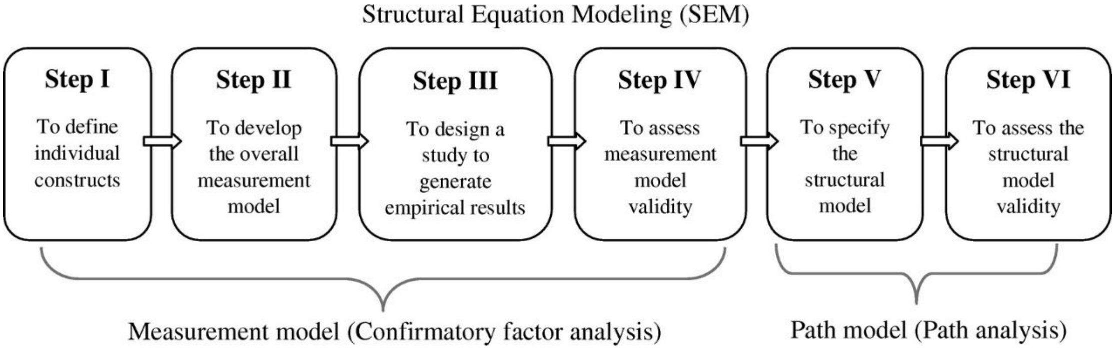

Structural Equation Modeling (SEM)
Step I
To define individual constructs
Step II
To develop the overall measurement model
Step III
To design a study to generate empirical results
Step IV
To assess measurement model validity
Step V
To specify the structural model
Step VI
To assess the structural model validity
Measurement model (Confirmatory factor analysis)
Path model (Path analysis)

Figure 16: Steps in Structural Equation Modeling Process (Hair et al., 2009)

Step I involves the definition of constructs. Hair et al. (2009) cautions that the Structural Equation Modeling process should not begin without first establishing a strong theoretical foundation for the measurement and structural models. The relationships specifying the model must be grounded in theory, particularly because Structural Equation Modeling is considered a confirmatory analysis technique that is guided more by theory than by empirical results. This research has undertaken a rigorous process in establishing a strong theoretical foundation for the multi-dimensional model as discussed in Chapter 2 and Chapter 3. The constructs chosen for this study have a theoretical underpinning and an empirical justification for their existence in the model.

Step I also involves operationalizing the constructs in the model by selecting measurement items and scales. This study has discussed this in detail in Section 4.7 for each of the variables in the model. The measurement items and scales have been informed from previous studies as summarized in Table 10. An essential part of this process is the pretest and pilot of the measurement instrument to ensure instrument validity which was done for this study as discussed in Section 4.8.

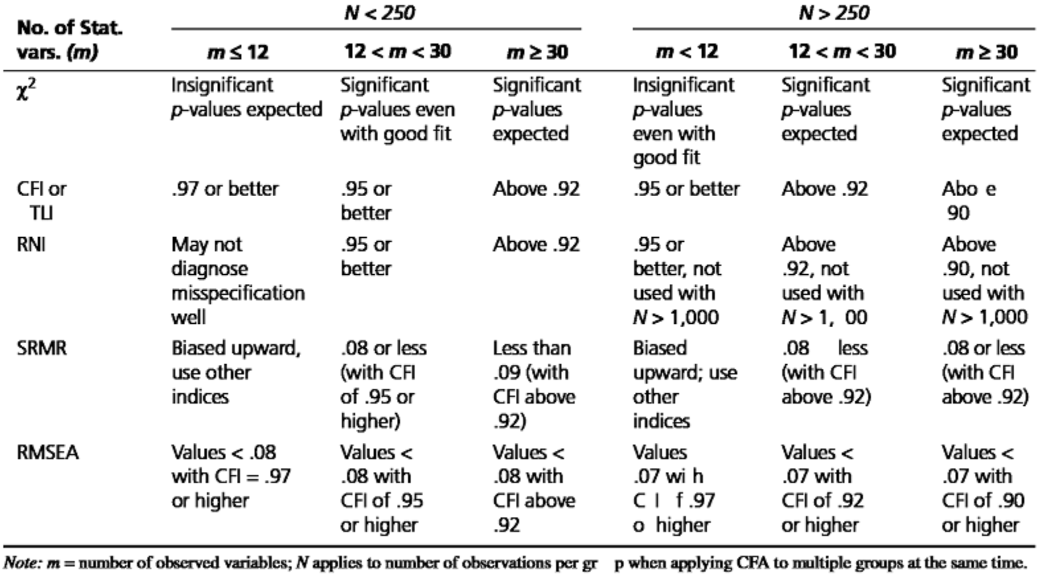
Step II is about development and specification of a measurement model. This step is often made simpler through the use of a path diagram. The path diagram illustrates the different variables and indicators that make up the measurement model. Key considerations at this step are how many indicators should be used for each construct and also whether the constructs are formative (the indicators are combined into an index) or reflective (the indicators are a result of the construct).

Step III is concerned with designing an appropriate study that can capture good quality data for testing the model. This chapter has outlined a clear research philosophy, design, research setting and study sample size suitable for testing the proposed model.

Step IV addresses measurement model validity which is achieved by attaining acceptable goodness-of-fit indices and establishing construct validity as outlined in Figure 17. Goodness-of-fit (GOF) checks how well the researcher’s theory compares to the reality observed in the data. According to Hair et al. (2009) these fit indices fall in three categories: (1) absolute fit indices (such as Chi-square test, Goodness of Fit Index (GFI) and Root Mean Square Error of Approximation (RMSEA)), (2) incremental fit indices (such as Adjusted Goodness of Fit Index (AGFI), Comparative Fit Index (CFI), Tucker-Lewis Index (TLI) and Normed Fit Index (NFI)) and (3) parsimony fit indices such as Parsimonious Goodness-of-fit Index (PGFI) and Parsimony Normed Fit Index (PNFI)).

Step V involves the specification of the structural model by specifying and assigning theorized relationships among constructs based on specified hypothesis.

Step VI is the final step that involves analyzing the validity of the specified structural model. This is done only after the measurement model is found satisfactorily valid. The overall fit of the structural model can be assessed using the criteria that was used for the measurement model. The closer the structural model fit is to the measurement model fit the better. In addition, each specific hypothesis needs to be tested to check whether the path estimates are significant and in the hypothesized direction. Finally, analysis of the variance explained by the model (R^2) should be performed.



No. of Stat. vars. (*m*)	*N* < 250			*N* > 250		
	m ≤ 12	12 < *m* < 30	*m* ≥ 30	*m* < 12	12 < *m* < 30	*m* ≥ 30
χ^2	Insignificant *p*-values expected	Significant *p*-values even with good fit	Significant *p*-values expected	Insignificant *p*-values even with good fit	Significant *p*-values expected	Significant *p*-values expected
CFI or TLI	.97 or better	.95 or better	Above .92	.95 or better	Above .92	Abo e 90
RNI	May not diagnose misspecification well	.95 or better	Above .92	.95 or better, not used with *N* > 1,000	Above .92, not used with *N* > 1, 00	Above .90, not used with *N* > 1,000
SRMR	Biased upward, use other indices	.08 or less (with CFI of .95 or higher)	Less than .09 (with CFI above .92)	Biased upward; use other indices	.08 less (with CFI above .92)	.08 or less (with CFI above .92)
RMSEA	Values < .08 with CFI = .97 or higher	Values < .08 with CFI of .95 or higher	Values < .08 with CFI above .92	Values .07 wi h C I f .97 o higher	Values < .07 with CFI of .92 or higher	Values < .07 with CFI of .90 or higher

Note: m = number of observed variables; *N* applies to number of observations per gr p when applying CFA to multiple groups at the same time.

Figure 17: Goodness-of-Fit Indices Across Different Models (Hair et al., 2009)

3.11.3 Analysis Procedures

This section outlines the various procedures undertaken to analyze the data collected in this research right from the beginning, including Structural Equation Modeling analysis to the final research findings.

1. ***Response Rate***

The first procedure is to establish the response rate in data collection. This is also sometimes referred to as the completion rate or the return rate. Baruch & Holtom

(2008) explain that it is important to examine whether any bias has been introduced due to non-response by a certain segment of the population or sample. They also explain that 100% response rate may not be possible where studies require voluntary participation. In their review of all articles published in 17 refereed journals in management and behavioural sciences they found that the average response rate was 52.6% across 152 studies with a standard deviation of 19.7. Their trend analysis from a previous study showed a significant decline in response rates over the years from 64.4% in 1975 to 48.4% in 1995.

Previous studies examining insider susceptibility to phishing have reported very low response rates simply due to the nature of the study. The study by Jagatic et al. (2007) reported a 16% phishing rate. The study by Mohebzada, El Zarka, Bhojani, & Darwish (2012) where two large-scale experiments were staged in a university community recorded a 8.74% rate for the first experiment and a 2.05% rate for the second experiment. A more recent Verizon (2018) investigation’s report data shows that 4% of those targeted in any phishing campaign will click on the phishing links.

It is important to re-emphasize that the focus of this study is to objectively identify insiders who are susceptible to unintentional threats and to study why they are susceptible to the threat. The segment of insiders susceptible to the threat may not be as large as the response rates found in pure survey studies as reported by Baruch & Holtom (2008). This is due to the additional staged experiment component that functions as an inclusion criterion to the subsequent questionnaire survey.

2. *Data Entry and Coding*

The next procedure after data collection is transcribing the data onto a computer file that can then be used for analysis on statistical software. Data entry and coding involves the use of unique identifiers for each questionnaire, variable labels that distinguish each variable and its indicators, definition of variable data and numerical codes to capture values for participant responses. Saunders et al. (2009) explains that the use of codes makes data entry much faster and the resulting file much easier to use in analysis. The underlying meaning of the codes is kept in a codebook to allow for reference and future decoding. This step essential for efficiencies in data analysis. In addition, Bhattacherjee (2012) points out that the data entry should be done into a file format that allow the data to be shared across different applications for statistical analysis. After the data is entered it should also be reviewed to ensure that no errors are introduced or omissions occur at this stage.

3. *Data screening*

The next procedure is data screening. This is done to examine the data for anomalies that may negatively impact the Structural Equation Modeling. This ensured that the data was of the right quality before further inferential analysis and modeling activities were undertaken. Hair et al. (2009) explains that this is a crucial analysis procedure that is often overlooked. The data should be screened for missing data, outliers, common method bias and assumptions for normality as explained hereafter.

i. *Missing Value Analysis*

Missing data can occur due to a failure of respondents to answer questions on the questionnaire but can also result from poor data entry. Hair et al. (2009) explain that missing values can be accommodated without outright removal if they are less than 10% of an examined case or 15% for a variable.

In addition, Hair et al. (2009) emphasize that analysis should be done to verify that there is no specific pattern associated with the missing values. It is important to establish whether the values are missing completely at random (MCAR). This allows for the missing values to be remedied by providing replacement values in a process referred to as imputation. Imputation provides data for missing values using existing valid values of the same case or from other cases in the sample.

ii. *Outlier Detection*

The next aspect of data screening is to identify outlier cases. Hair et al. (2009) defines outliers as cases that have unique characteristics that distinctly set them apart from other observations in the data set. Outliers usually have very high or very low values that distinguishes them from the rest of the data.

Outliers can be problematic in statistical analysis because they may distort statistical tests and analysis because they are not a good representation of the population and can be considered as influential cases (Field, 2009). A caution is however not to classify all outliers as problematic, they need to be evaluated in the context of the study being carried out (Hair et al., 2009). Although they may be markedly different from the rest of the population, they may point out characteristics that would not have been seen with normal cases. Novel discoveries could be made when analyzing outliers.

There are two main methods for outlier detection as outlined by Hair et al. (2009): univariate outlier detection and multivariate outlier detection.

In univariate outlier detection, the distribution of values for each variable is examined and outliers are marked as the cases that are higher or lower than the defined threshold ranges of the distribution. Field (2009) explains that if we consider a normally distributed dataset, we expect 5% of the cases to be greater than the absolute value of 1.96, 1% to be greater than the absolute value of 2.58 and none to be greater than the absolute value of 3.29. This study therefore considers an outlier as any value that is greater than the ± 3.29 threshold for standardized Z-scores.

The second method is the multivariate outlier detection which tries to identify influential cases that may have an impact on the model parameters. The Mahalanobis D^2 measure is a recommended technique which provides a probability statistic based on Chi-square Cumulative Distribution function. Any case with a statistically significant statistic, where the probability value is $p < 0.001$, is considered an outlier.

It must be noted that Field (2009) and Hair et al. (2009) do not recommend outright removal of outliers. They should only be removed after reviewing the cases and after it is clear that they are not a representation of any possible segment of the population. If they are removed without careful review the results of the study may be argued not to be generalizable.

iii. Common Method Bias

The use of the questionnaire as a survey instrument for data collection has been associated with many types of bias and one of the key ones is the Common Method Bias (CMB). Bhattacherjee (2012) explains that the CMB is the covariance between independent and dependent variables that results from having been measured using the same survey instrument at the same time. The covariance is therefore introduced by the instrument and measurement method and not because it exists in the phenomenon.

To counter common method bias in this research, two techniques are used to measure the dependent variables to address single-rating issues. One uses directly observed behaviour and the other uses self-reported questionnaire items. Both measures are integrated in the data set before the analysis procedures were started.

In addition, the data set is tested for the presence of common method bias using Harman's one-factor test using guidelines from (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003). Using this test, CMB is likely to exist if measurement items are found to load on a single factor. In addition, this is further confirmed if this single factor is found to explain more than 50% of the variance on all items.

iv. Normality

Normality is one of the key assumptions for many different statistical analysis techniques. Normality assesses the extent to which variable data distributions conform to the symmetrical bell-shaped curve associated with a random variable which has a kurtosis of 0 and a skew of 0 (Field, 2009).

Visual methods using normal curves on histograms, box plots, Q-Q plots and P-P plots can be sued to visually determine if data is normally distributed. The other method for determining normality is through the use of normality tests such as the Kolmogorov-Smirnov (K-S) and Shapiro-Wilk tests. Another way of determining normality is by assessing the skewness and kurtosis values for variables as a univariate method. Curran, West, & Finch (1996) recommends that skewness values be ≤ 2 and kurtosis values be ≤ 7 as a sign of normality.

4. Descriptive Analysis

The next analysis procedure is the general exploration of the data and description of its characteristics using descriptive statistics. Common descriptive statistics include measures of central tendency (such as the mean, median and mode) and, measures of dispersion (such as the range and standard deviation). These provide a way for statistically describing the data in meaningful ways.

5. Development of the Measurement Model

Structural Equation Modeling is primarily a combination of factor analysis and path analysis (Weston & Gore, 2006). The factor analysis component is concerned with developing the measurement model. The measurement model examines how well the measured (observed or manifest) indicators derived from the measurement tool (such as a questionnaire) combine to form their target unobserved (latent) factors.

Brown (2006) explains that the primary aim of factor analysis is to discover the number and nature of latent factors that account for the correlations among the observed indicators. Factor analysis is based on the premise that the observed indicators are correlated because they share a common factor. The two forms of factor analysis are Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA) and these are compared in Table 11.

Table 11: Comparison of Exploratory and Confirmatory Factor Analysis (Hassan & Abu Bakar, 2009)

Similarities	Differences
• Their aim is to reduce a large number of indicators to a smaller set of factors • Factors are defined from variables that are highly correlated to each other • A theory or conceptual background is needed to explain the factor structure and the relationships • Both can be done by IBM SPSS. EFA is done on SPSS Statistics and CFA is done on SPSS AMOS	• EFA does not start with a pre-defined number of factors or a pre-defined indicator-factor relationship. However, CFA starts with a theorized indicator-factor structure with a fixed number of factors defined from theory. • EFA is useful in exploratory theory building since the factor structure is generated from the data. CFA is used for confirmatory theory testing since a predefined factor structure is used in the analysis. • EFA is suited for testing new or modified measurement scales. However, CFA takes already established measurement scales and replicates construct validity to other samples.

Both Exploratory and Confirmatory Factor Analysis are done in this research. Exploratory Factor Analysis is particularly necessary because some measurement scales are modified for this research. It will be important to see if the measured indicators conform to the expected factor structures. The Confirmatory Factor Analysis is a critical step in Structural Equation Modeling and will be necessary for testing the validity and goodness-of-fit of the proposed multi-dimensional model for determining susceptibility to Unintentional insider threats.

6. *Exploratory Factor Analysis*

The Exploratory Factor Analysis is done to condense a large number of highly correlated indicators to smaller number of factors. The measurement model that EFA estimates is not restricted by a pre-defined or theorized indicator-factor relationship. The associations are synthesized from the data itself during analysis allowing for an exploratory generation of new theories (Hassan & Abu Bakar, 2009). EFA is done before CFA because it targets scale development and construct definition. EFA is also very good at testing new or modified scales to see whether they measure constructs well. EFA checks whether the underlying constructs that were targeted are actually unveiled from the data (Brown, 2006).

IBM SPSS Statistics version 23 is used to conduct an Exploratory Factor Analysis on the research data. The Maximum Likelihood factor extraction method is used because it provides a goodness of fit evaluation for the factor selection to get the most appropriate measurement model. The promax oblique method is used for factor rotation in order to obtain a simplified factor structure. The pattern matrix is then refined iteratively in order to eliminate poorly behaved indicators, particularly those that had very small loadings (also called low communalities) and those that had high loading on more than one factor (also called cross loading). After eliminating poor indicators, the EFA is re-run until a good factor structure is obtained.

7. *Confirmatory Factor Analysis*

Confirmatory Factor Analysis is similar to Exploratory Factor Analysis in that it also aims to describe the relationships between a group of indicators and a smaller set of latent variables. However, unlike Exploratory Factor Analysis, the Confirmatory Factor Analysis process starts from a priori theorized relationship between indicators and factors. A prior conceptual or empirical foundation needs to be developed to guide the specification and testing of the confirmatory factor model. The researcher needs to clearly define the number of factors, the pattern of indicator-factors and their loadings – typically after conducting an Exploratory Factor Analysis (Brown, 2006).

The CFA and EFA in this research both used the Maximum Likelihood method for factor extraction and estimation. This is because the Maximum Likelihood method provides a rich set of indices for evaluating the appropriateness and goodness-of-fit of the proposed factor solution. The provided CFA factor solution provides input for specification of the structural model which defines how various factor are related to each other based on a determined theoretical foundation. This can be simplified in what is called a path diagram.

IBM SPSS AMOS Graphics version 23 is used for the Confirmatory Factor Analysis and also for the conversion of the resulting measurement model into a structural model for further analysis.

8. *Validating the Measurement Model*

One of the key objectives of Confirmatory Factor Analysis is the validation of the measurement model (Awang, 2012; Hassan & Abu Bakar, 2009). The CFA allows for each factor to be validated with respect to its unidimensionality, reliability and validity. Any item or factor that does not pass the CFA should be dropped.

i. *Unidimensionality*

Awang (2012) explains that the unidimensionality criteria is met if all measurement items have satisfactory loadings for their respective latent factor. Items that have very low factor loadings should be dropped. Hair et al. (2009) states that any factor loading less than 0.5 are considered low and in fact all factor loadings defining a particular latent factor should average above 0.7. Deletion of items should be done one at a time with the lowest loading items being dropped first. The analysis is ran again after deletion of an item. This continues until the unidimensional criteria is met for all factors. However, the caution is that no more than 20% of items should be deleted. In addition, Hair et al. (2009) explains that standardized loadings should not be higher than +1.0 or lower than -1.0 otherwise they would indicate a problem with the data.

ii. *Reliability*

Reliability assesses the extent to which a set of measurement items can be treated as measuring a single latent construct because they are highly correlated. This refers to the internal consistency of the items in a summated scale. Hair et al. (2009) explain that Cronbach’s Alpha is the most widely used measure of reliability and its value should be greater than 0.70 although for exploratory research it can be allowed to go as low as 0.60. The other reliability measure that is mostly used with Confirmatory Factor Analysis is the Composite Reliability (CR). Nunnally & Bernstein (1994) prescribe that the CR should be at least 0.70 for developing instruments but can be required to be at least 0.80 for advanced stages of instrument development.

iii. Validity

Validity checks whether the instrument items and scales measured what they were supposed to measure for a latent construct. Awang (2012) explains that three types of validity should be examined for each construct: construct, convergent and discriminant validity.

Construct validity examines the extent to which the set of measured items reflects the theoretical latent construct they are meant to measure. Hair et al. (2009) explains that convergent validity is determined by the convergent, discriminant, nomological and face validity. Face validity is determined before the administration of the measurement instrument. Nomological validity is determined by examining correlations among constructs. However, Awang (2012) explains that construct validity is determined by checking if the various fit indices are at a satisfactory level. These fit indices fall in three categories: (1) absolute fit indices (such as Chi-square test, Goodness of Fit Index (GFI) and Root Mean Square Error of Approximation (RMSEA)), (2) incremental fit indices (such as Adjusted Goodness of Fit Index (AGFI), Comparative Fit Index (CFI), Tucker-Lewis Index (TLI) and Normed Fit Index (NFI)) and (3) parsimony fit indices such as Parsimonious Goodness-of-fit Index (PGFI) and Parsimony Normed Fit Index (PNFI)).

The convergent validity requirement is met when all items that make up the measurement model are found to be statistically significant. The convergent validity is verified using the Average Variance Extracted (AVE) statistic that is calculated for each construct. Awang (2012) explains that the AVE should be at least 0.5 and higher because this means at least 50% of the variance in indicators is accounted for by the latent construct and not by measurement error. Low factor loading items are known to negatively affect the convergent validity of a latent construct.

Discriminant validity is the extent which a latent construct is distinct and different from all the other latent constructs in the model. In fact a construct must not be highly correlated with another factor. Brown (2006) states that an inter-factor correlation of .80 and above is an indication of poor discriminant validity and advice that such constructs should be collapsed together into a single construct. This may, however, degrade the model fit. When a construct achieves discriminant validity it means that it is able to account for more variance in the observed items associated with it than with other constructs in the model. Fornell & Larcker (1981) present another method for assessing discriminant validity using the Average Variance Extracted (AVE). They prescribe that the AVE for any two constructs need to be greater than their shared variance. This means that the square root of AVE for every construct should be greater than any correlation among any pairs of constructs.

Table 12 summarizes the various criteria for evaluating the measurement models before proceeding to the structural model.

Table 12: Summary of Evaluation Criteria for Measurement Models

Criteria	**Measure**	**Desired Threshold Values**
Internal Consistency	Cronbach's Alpha	≥ 0.70 for initial stages of measurement development ≥ 0.80 for advanced stages of measure development (Nunnally & Bernstein, 1994)
Convergent Validity	Composite Reliability	≥ 0.70 for adapted instruments ≥ 0.80 for advanced instrument development (Nunnally & Bernstein, 1994)
	Average Variance Extracted (AVE)	≥ 0.50 (Fornell & Larcker, 1981; Hair et al., 2009)
Discriminant Validity	$\sqrt{\text{AVE}}$	The square root of AVE for a construct should be greater than its correlation with any other factor (Fornell & Larcker, 1981)
	Cross loadings	The correlation of indicators with their associated constructs should be higher than with any other construct (Brown, 2006)

9. *Development of the Structural Model*

Hair et al. (2009) points out that the Confirmatory Factor Analysis provides the foundation for the theoretical testing of relationships between latent constructs in the structural model stage. A key strength of Structural Equation Modeling is that it takes in the measurement model into account when testing the structural model. The structural model uses a diagram to represent theory and depict relationships between the constructs.

i. *Construction of Path Diagram*

The structural model is developed from the measurement model by modifying its path diagram. Hair et al. (2009) explains that the two-headed arrows between factors are converted to single-headed arrows representing a theorized cause-effect relationship as illustrated in Figure 18 from Brown (2006). The first path diagram shows the resulting measurement model after the Confirmatory Factor Analysis process and the second diagram shows how the measurement model is modified into a structural model that specifies relationships between the factors.

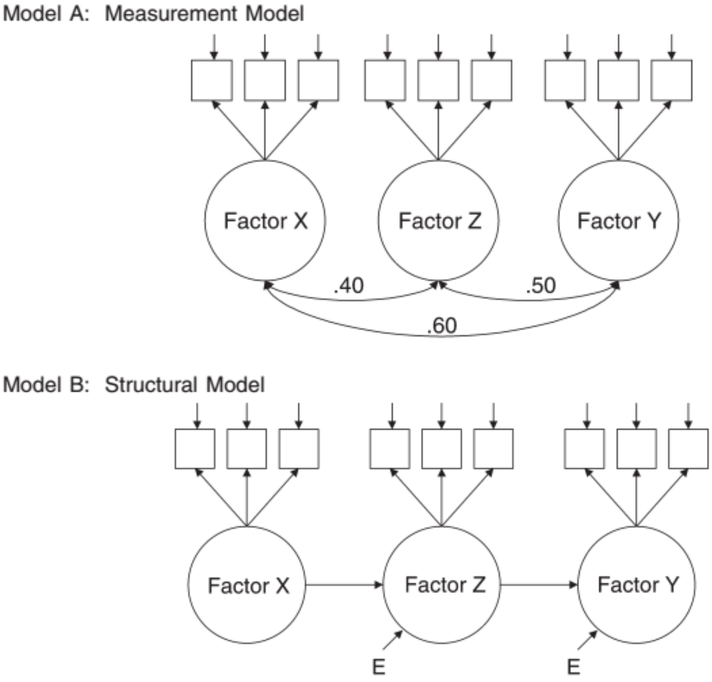


Figure 18: Path Diagram of Measurement and Structural Models (Brown, 2006)

10. Validating the Structural Model

The assessment of the structural model has to take into account the model fit and also the analysis the parameter estimates on the path diagram.

i. Model Fit

The measurement model fit is evaluated during the Confirmatory Factor Analysis process using the same indices that are used to assess the model fit of the structural model. Hair et al. (2009) point out that the structural model is not expected to have a better model fit than the measurement model. In fact, the structural model process does not improve the model fit. This means that the adequate model fit has to be achieved for the measurement model before moving to the structural model.

The structural model seeks to define the most significant relationships between constructs and is therefore expected to be a simplified and precise explanation of the model as compared to the measurement model used in Confirmatory Factor Analysis.

There are various model fit indices to be considered and according to Hair et al. (2009), they can be categorized into 3 general groups: absolute, incremental and parsimony fit measures. Table 13 mentions the common indices for each category and their desired value ranges.

Table 13: Goodness-of-Fit Indices and Their Desired Thresholds (Hair et al., 2009)

Goodness-of-Fit Measure	Notation	Desired Threshold
Absolute Measures		
Chi-square test	χ^2	$p > .05$
Degrees of freedom	df	≥ 0
Chi-square/df ratio	χ^2/df	< 3
Goodness of Fit Index	GFI	$> .90$
Root Mean Square Error of Approximation	RMSEA	$< .80$
Incremental Measures		
Adjusted Goodness of Fit Index	AGFI	$> .90$
Tucker-Lewis Index	TLI	$> .90$
Normed Fit Index	NFI	$> .90$
Comparative Fit Index	CFI	$> .90$
Parsimonious Fit Measures		
Parsimony Normed Fit Index	PNFI	$> .50$
Parsimonious Goodness-of-fit Index	PGFI	$> .50$

Absolute measures provide the most basic evaluation of how well the theoretical model fits the sample data. The Chi-square (χ^2) statistic is the most fundamental fit measure for Structural Equation Modeling. However good χ^2 is difficult to achieve for models with many indicators and large sample size. It is therefore not used alone. Normed Chi-square is a ratio of χ^2 to degrees of freedom (*df*) and is also used as a measure of fit. Ratios of 3:1 or less are associated with good fit. The Goodness-of-Fit Index (GFI) is the other statistic that is used. GFI has an advantage over χ^2 because it is affected less by sample size. GFI values are expressed between 0 and 1 and the higher the value. Values greater than 0.90 are considered good. Root Mean Square Error of Approximation (RMSEA) is even more preferred over χ^2 because it corrects for model complexity and sample size. Low RMSEA values within the range of 0.03 to 0.08 are considered good. Root Mean Square Residual (RMR) and Standardized Root Mean Residual (SRMR) are the other indices where low values represent better fit than higher values. For this reason, they are sometimes referred to as badness-of-fit measures.

Incremental measures assess how well a model compares to an alternative baseline model such as the null model. The null model assumes that all the observed variables are uncorrelated. A SEM analysis tool may not generate all of them. The Normed Fit Index (NFI) is one of the first incremental indices but is less commonly used. NFI values range between 0 and 1 where perfect fit would be a value of 1. Comparative Fit Index (CFI) is an improved version of the NFI and is one of the commonly used incremental index. CFI values greater than 0.90 are considered good fit. Tucker Lewis Index (TLI) is another incremental index that is preferred over NFI because it takes model complexity into account. Values do not fall between 0 and 1 because it is not normed. However, values that approach 1 are considered good fit. Relative Noncentrality Index (RNI) is another incremental measure whose values are normed between 0 and 1. RNI values greater than 0.90 are associated with good fit.

Parsimony fit measures consider fit relative to model complexity. The measures are improved either by achieving better fit or having a simpler model with fewer estimated parameter paths. Parsimony fit indices are more useful when used to compare competing models. Adjusted Goodness of Fit Index (AGFI) adjusts GFI by a ratio of the degrees of freedom. It is however rarely used because it is affected by model complexity and sample size. Parsimony Normed Fit Index (PNFI) is the more widely used parsimony fit measure.

ii. Hypothesis Testing

Structural model evaluation also involves the examination of various parameter estimates that indicate whether hypotheses hold or not. Hair et al. (2009) explains that one of the things to examine is if the path coefficients (also called β values) are in the predicted direction. Parameters that are greater than 0 indicate a positive relationship between constructs while those less than 0 indicate a negative relationship. Analysis should also be done to determine if paths are significant at either 0.1, 0.05 or 0.001 level by examining their associated p-values. The other key analysis is regarding the coefficient of determination (R^2) which represents the proportion of the endogenous construct's variance that is explained by its predictors. Table 14 summarizes the various criteria for evaluating structural models.

Table 14: Summary of Evaluation Criteria for Structural Models (Hair et al., 2009)

Criteria	**Desired Threshold Values**
Path Coefficients (β) signs and significance	• Values close to ±1 suggest a strong relationship or influence between variables • Values close to 0 indicate a weak relationship • Values higher than ±2 can be considered substantial
Coefficient of Determination ($\boldsymbol{R^2}$)	• Gauges predictive power that predictor variables have on dependent variables • Values should be a minimum of 0.10 • High values are required for the model to be considered to have significant explanatory power

3.12 Ethical Considerations

As pointed out by Finn & Jakobsson (2007), there are various ethical issues to be addressed when conducting information security research, particularly when targeting naturalistic observation of unintentional insider threat behaviour.

Research ethics relates to moral choices and decision making relating to research conduct (Greener, 2008). Various principles have to be upheld in the course of a research. These include honesty, integrity, objectivity, respect for intellectual property, confidentiality and protection of research participants. Diener & Crandall (1978) highlight four main issues relating to research ethics: harm to participants, deception, invasion of privacy and lack of informed consent.

This research took special care to address these ethical concerns. Institutional approval to conduct research at the selected site was obtained from its research office and technology department. This provided a site approval to conduct the research and collect the data from the organization’s insiders. In addition, an Institutional Research Board (IRB) approval of the research proposal, including data collection procedures and methodologies, was obtained. The IRB approval signified that the research was found to meet the required standards and was not going to be harmful to the participants or the organization.

A key concern for the IRB as pointed out by Finn & Jakobsson (2007), was the use of deception in order to observe natural insider behaviour. The study participants were not alerted about the phishing exercise in order to avoid the Hawthorne effect (Parsons, 1974). Participants are known to change their behaviour when they know they are being studied. The naturalistic exercise aimed to mimic a real-world social engineering threat in every way and to observe the study participant reactions to the attack. This is the recommended method of studying Unintentional insider threats because it provides non-biased objective results (Huber et al., 2009; Kumaraguru et al., 2009, 2008; Workman, 2007, 2008a).

The role of the IRB is to ensure that this naturalistic study does not pose any actual harm to the participants or the organization. To ensure this, two senior ICT administrators were attached to the research to review the phishing email and website to ensure that none of the technical components harmed the organization's information system or collected sensitive data from the insiders.

In addition, informed consent of the participants would be obtained before administering questionnaires for data collection. This meant that the sampled participants would be given an overview of the research and would need to give their approval for data to be collected about them or the already collected data to be used in the research. The participants were allowed to withdraw their participation at this time or any other time they desired to. This provided another key element of research ethics which is voluntary participation.

Diener & Crandall (1978) differentiate confidentiality and privacy and emphasize the need for research to fulfill these two key ethical considerations. Confidentiality is upheld in all stages of the research by making sure that: study participants are anonymized and no data is personally identifiable to them. Privacy regards the usage of the research data and this study ensures that the detailed raw data is not disclosed to other entities other than the researcher and the appointed academic supervisory teams. In addition, and published results are reported in collective terms where the organization or study participants are not identifiable.

3.13 Chapter Summary

This chapter has extensively outlined this study's research philosophy, strategy and design. This research has taken up a realist ontological view, positivist and objective epistemological philosophical stance. A clear justification for this chosen research philosophy has been given with the main reasons being the well-defined guidelines for research, scientific approach relating to theory, discovery of cause-effect relationships and provision of a robust foundation for future research that provides a multi-dimensional model for determining susceptibility to Unintentional insider threats. A deductive research design that employs the use of cross-sectional data captured using a questionnaire survey was selected and justified for this study. This research design is the recommended approach for such research because it allows for quantitative data to be collected in testing a theoretical model.

This chapter has also provided an explanation for the selection of the research site and setting which is a private university located in Nairobi, Kenya. Key elements for the selection decision are the ability to get research approvals and cooperation to conduct a naturalistic field study that provides the best ecological validity for studying the Unintentional Insider Threat phenomenon. Details of the research population, sampling frame, sampling technique and resulting study sample have been provided with justifications given based on the ability to obtain probabilistic results that can be generalized to a particular population.

Data collection procedures that take place through observations in a naturalistic field study and the use of a questionnaire survey have also been discussed. The development and validation of the data collection instruments has been discussed to a great extent and measurement scales have been presented from extant literature.

In addition, this chapter covers the data analysis procedures in great detail. It outlines the steps to be followed which include, determination of the response rate, data entry and coding, data screening (for missing values, outliers, common method bias and normality), descriptive analysis and inferential analysis based on the Structural Equation Modeling process (measurement model development and validation, exploratory and confirmatory factor analysis, structural model development and validation). Various ethical issues that were considered and designed into the research process are also presented with the key objectives of obtaining institutional research approvals and informed consent; protecting participants from harm; providing confidentiality and privacy of research data and results.

CHAPTER 4: RESULTS

4.1 Introduction

This chapter provides the results of the analysis procedures applied to the data collected in this research. Specifically, it outlines the results of the analysis of the response rate, data screening (for missing values, outliers, common method bias and normality), descriptive statistics of study participants, exploratory factor analysis, confirmatory factor analysis, measurement model validation, structural model validation and hypothesis testing. Licensed versions of IBM SPSS Statistics version 23 and IBM SPSS AMOS version 23 applications were used in the various analysis procedures and in Structural Equation Modeling.

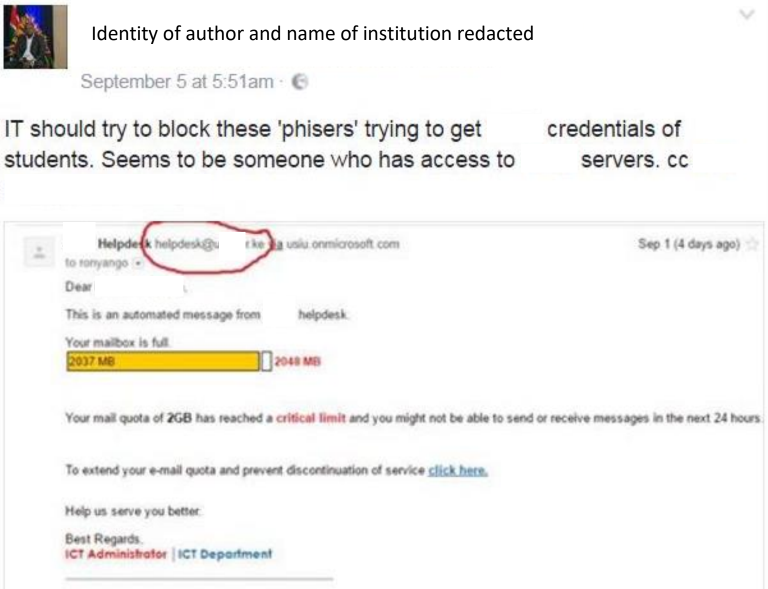
The general objective of this research was to develop and validate a unified multi-dimensional theoretical model for determining susceptibility to the Unintentional Insider Threat to information systems security. The various factors that contribute to the Unintentional Insider Threat were explored from extant literature as presented in Chapter 2 thereby fulfilling the first specific objective of this research. The second specific objective of this research was fulfilled in Chapter 3 whereby a unified multi-dimensional theoretical model that explains why people are susceptible to the Unintentional Insider Threat was presented. This chapter fulfils the third specific objective of this research which is to validate this model using empirical data and appropriate statistical methods.

4.2 Response Rate

Data for this study was collected through two activities as outlined in Section 3.5. The first was through a staged naturalistic phishing exercise which simulated a real-world attack towards a sampled university population. The data collected at this stage provided an objective measure of the unintentional insider threat outcome behaviour. The second data collection was through self-reported questionnaires.

4.2.1 Responses from Naturalistic field study

The phishing exercise started on 28th July 2016 and ran till 6th September 2016. It should be noted that the study was stopped at this time because it drew negative publicity within the university community. On the 5th of September a prominent social media activist and blogger who was a student at the university posted a comment on the university's social media pages alerting the community of the phishing email and calling it to be investigated and stopped. The social media post is shown in Figure 19. This prompted the executives at the university to call off the exercise due to the negative image the social media post had painted. The ICT director, who had been involved in the research approvals and was aware of its progress, instructed his team to send out alerts to the entire university community informing them of the nature of the research and allaying any concerns that had been raised by the social media post.



Identity of author and name of institution redacted

September 5 at 5:51am ·

IT should try to block these 'phisers' trying to get [redacted] credentials of students. Seems to be someone who has access to [redacted] servers. cc

Helpdesk helpdesk@u[redacted]r.ke via usiu.onmicrosoft.com Sep 1 (4 days ago)
to ronyango

Dear [redacted]

This is an automated message from [redacted] helpdesk.

Your mailbox is full.
2037 MB 2048 MB

Your mail quota of **2GB** has reached a critical limit and you might not be able to send or receive messages in the next 24 hours.

To extend your e-mail quota and prevent discontinuation of service click here.

Help us serve you better.

Best Regards,
ICT Administrator | ICT Department

Figure 19: Phishing Alert by Prominent Blogger

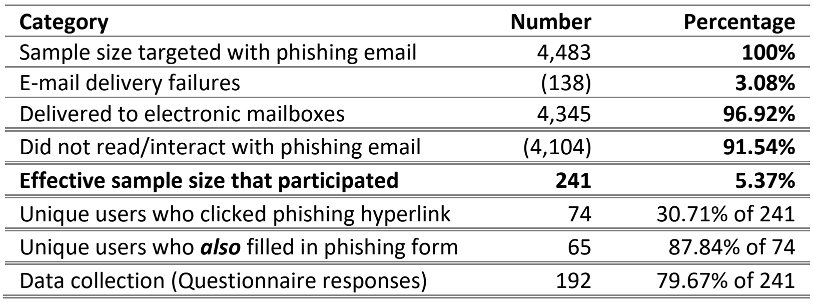
The staged phishing exercise ran for 40 days and within this time all the 4,483 insiders who were sampled from the university community had already been sent phishing emails through their official university accounts. The email system returned delivery failures for 138 of the emails indicating that there was a problem with the email account. This meant that 4,345 phishing emails were actually delivered to the insiders' official email accounts. The research involved the purchase of an or.ke domain mimicking the institution's .ac.ke domain. It also involved renting both a mail and web server for hosting the phishing emails and phishing website. The data regarding which insiders were susceptible to the attack was obtained from the hosting servers. These servers had records of the individuals who opened emails, clicked hyperlinks and filled in form data. The data indicated a total of 98 clicks from the phishing emails. These clicks were associated with 75 unique insiders since some clicked the phishing hyperlink multiple times. In addition, the form on the phishing website was filled in 72 times with 66 form-fills being unique and the others being repeated entries.

Statistics on interaction with the phishing email were low. There was no response or interaction with the phishing email by 4,104 (91.54%) of the targeted sample. The emails were sent to the insiders using their official university accounts. Discussions with the ICT staff attached to the study revealed that it could be that few people used their official university accounts for correspondence. Students (who were the largest number in the sample) had an option of using alternative email addresses to receive communication from the university. This meant that they had no imperative to use their official email accounts. Instead they preferred to use private email accounts mainly from Google, Hotmail or Yahoo.

4.2.2 Responses from Self-reported Questionnaire Survey

The questionnaire was then administered to all 241 study participants who were identified as having interacted with the phishing instruments. These are the people who received the phishing emails and actually opened them. Read receipts were setup in Microsoft Outlook to give this indication. These 241 individuals can be argued to be the effective sample size in this research study and represents 5.37% of the total sample size as outlined in Table 15. Interaction with the staged phishing attack was considered as an inclusion criterion for the next step of data collection using the questionnaire survey. Of the 241 possible respondents, 192 filled in and returned their questionnaires. The effective response rate from the administered questionnaires was therefore 79.67%.

Table 15: Response Rate Statistics



Category	Number	Percentage
Sample size targeted with phishing email	4,483	**100%**
E-mail delivery failures	(138)	**3.08%**
Delivered to electronic mailboxes	4,345	**96.92%**
Did not read/interact with phishing email	(4,104)	**91.54%**
Effective sample size that participated	**241**	**5.37%**
Unique users who clicked phishing hyperlink	74	30.71% of 241
Unique users who ***also*** filled in phishing form	65	87.84% of 74
Data collection (Questionnaire responses)	192	79.67% of 241

Data collected on the backend database showed that there were a total of 98 clicks on the phishing hyperlink. These clicks were associated with 74 unique insiders since some clicked the phishing hyperlink multiple times, as indicated by repeated entries in the backend database. In addition, the form on the phishing website was filled in 72 times with 65 form-fills being unique and others being repeated entries. This shows that 87.84% of the insiders who were susceptible to phishing emails went ahead to disclose passwords that would enable an attacker gain access to the organization's systems.

The response rate per strata is provided in Table 16. Results shows that interns (25%), staff (22.89%), full-time faculty (17.33%) and mailing list users (14.29%) had higher response rates in proportion to the numbers targeted per strata. Students had a very low percentage (0.49%) of successfully phished per strata despite having the highest number in sample.

Table 16: Response Rate per Strata

Strata	Size in Sample	Successfully Phished	Proportion
Students	4,122	20	0.48 %
	166	38	22.89%
Adjunct Faculty	84	1	1.19%
Full-time Faculty	75	13	17.33%
Management	6	0	0%
Interns	4	1	25%
Mailing List Users	7	1	14.29%
Unknown	19	0	0%
Total	**4483**	**74**	

A recent study by Mohebzada, El Zarka, BHojani, & Darwish (2012) where two large-scale phishing exercises were conducted, only a 8.74% rate was achieved in the first experiment and a 2.05% rate in the second experiment. Additionally, a recent data breach investigation by Verizon (2018) establishes a 4% rate for phishing campaigns. Therefore, it can be argued that this study's response rate is within the expected range for a naturalistic field study. Although the phishing rates were low, it should be understood that it only takes a few users to compromise an information system. Once an attacker is successful with some systems, these can then be used as a pivot point to work into the rest of the organization (Ali, 2015).

4.3 Data Screening

The collected data was screened for missing values, outliers, common method bias and assumptions of normality as outlined by Hair et al. (2009). This was done in order to identify any concerns that could impact further analysis or even the Structural Equation Modeling process.

4.3.1 Missing Value Analysis

IBM SPSS Statistics software was used to examine the data for missing values. Of the 192 cases, 21 had missing data and the highest percentage of missing data was 8.2% for a case as outlined in Table 17. In addition, missing data was also analyzed per variable and the highest percentage of missing data was 1.6% per variable as outlined in Table 19. The actual physical questionnaires were reviewed to ensure that the missing values were not a result of poor data entry.

Since the missing data was below the 10% threshold for individual cases and 15% for variables, this did not warrant outright deletion of affected cases or variables. Imputation for missing values was done and replacement values were calculated.

Table 17: Missing Data per Case

PNO	Missing Count	Missing %
6	1	0.7
11	1	0.7
31	1	0.7
83	1	0.7
34	1	0.7
62	1	0.7
76	1	0.7
81	1	0.7
114	1	0.7
126	1	0.7
155	1	0.7
219	1	0.7
202	1	0.7
30	2	1.4
212	2	1.4
80	3	2.1
44	3	2.1
45	4	2.7
54	6	4.1
144	6	4.1
35	12	8.2

The missing values were replaced using the mean for ratio data and using the median for ordinal data to prevent undue influence of missing values on the analysis. This is illustrated in Table 18.

Table 18: Missing Data Imputation per Variable

Variables	Missing Count	Missing %	Imputation Replacement By
AGE	2	1.1	Median
EDUCATION	1	.5	Median
YEARS_UNI	2	1.1	Mean
YEAR_INTERNET	1	.5	Median
HOURS_INTERNET	1	.5	Median
ER1	1	.5	Median
OSI	1	.5	Median
OS2	2	1.1	Median
OS3	2	1.1	Median
OS4	1	.5	Median
RP1	1	.5	Median
RP2	2	1.1	Median
QSR_OB1	2	1.1	Median
QSR_OB2	2	1.1	Median
QSR_OB3	2	1.1	Median
BI1	1	.5	Median
BI2	2	1.1	Median
OF2	3	1.6	Median
TC1	1	.5	Median
TC2	1	.5	Median
TC3	1	.5	Median
SM1	1	.5	Median
PV1	1	.5	Median
PV5	1	.5	Median
PVUL2	1	.5	Median
PVUL3	1	.5	Median
PS1	1	.5	Median
PS3	1	.5	Median
PS4	1	.5	Median
PS5	1	.5	Median
PS6	1	.5	Median
DC1	1	.5	Median
DC2	1	.5	Median
DC3	1	.5	Median
PC2	1	.5	Median
MP1	1	.5	Median
MP2	1	.5	Median
MP3	1	.5	Median
AP3	1	.5	Median
DIST2	1	.5	Median

4.3.2 Outlier Detection

Outliers are cases with distinctly unique characteristics that set them apart from other observations in the data set. They usually have very high or very low values that may distort results of statistical analysis. They are also not a good representation of the population under study.

The multivariate outlier detection technique using the Mahalanobis D^2 measure was used to detect outliers for this study. Using IBM SPSS Statistics the Mahalanobis statistic was computed for each case in the data set and stored in a new variable (MAH_1). Thereafter a Chi-Square Cumulative Distribution Function was used to calculate the p-values. Any case with a p-value less than 0.001 would be considered an outlier. None of the p-values was less than 0.001 therefore it is concluded that there are no outlier cases in the data set.

4.3.3 Common Method Bias

The Common Method Bias is the covariance that exists in a dataset because it was introduced by the measurement method and not because it exists in the phenomenon under study (Bhattacherjee, 2012).

IBM SPSS Statistics was used to perform the Harman one-factor test where an un-rotated factor analysis was done with the aim of extracting only one factor. The resulting factor was found to explain only 17.13% of the total variance. This was much less than the cutoff threshold of 50% set by Podsakoff et al. (2003). It was therefore concluded that Common Method Bias was not a threat in the data set.

4.3.4 Normality

Questionnaire items based on Likert (1932) type measures are ideally thought of as ordinal measurement scales. Boone & Boone (2012) explain that these Likert-type measures need to be converted to composite scales in order to allow proper analysis. Following this recommendations, composite scales were created by composing summative scores from multiple questionnaire items measuring the same variable. These composite scales were then evaluated for normality.

Normality was analyzed by examining the skewness and kurtosis values for each measure. The guideline for test of normality from Curran et al. (1996) was to identify measures whose skewness values are greater than 2 and kurtosis values are greater than 3. Only four variables had normality issues as listed in Table 19.

Table 19: Normality Statistics

	Item	N	Skewness	Kurtosis
1.	RE	192	-1.622	3.443
2.	RC	192	1.845	3.518
3.	PB	192	-3.461	17.381
4.	PS	192	-2.264	6.604

Instead of immediately dropping the variables, it was decided to take note of these constructs and to examine them again particularly in the Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA).

4.4 Descriptive Analysis

The next analysis procedure involved conducting a general exploration of the dataset; primarily by using descriptive statistics. The demographic variables were targeted in this analysis and each is discussed in the following sections:

4.4.1 Gender

An analysis of the gender distribution showed that the majority of the respondents were male (63%). Only 37% were female as outlined in Table 20.

Table 20: Respondent Gender Distribution

Gender	Frequency N=192	Percentage
1. Male	121	63%
2. Female	71	37%

4.4.2 Age

The respondents were asked to indicate which age group they were in. Majority were in the 26 – 35 years age bracket while the least were above 55 years. This is summarized in Table 21.

Table 21: Respondent Age Distribution

Age in Years	Frequency N=192	Percentage
1. 18-25 years	45	23.4%
2. 26-35 years	66	34.4%
3. 36-45 years	44	22.9%
4. 46-55 years	20	10.4%
5. Above 55 years	17	8.9%

4.4.3 Level of Education

The respondents were asked to indicate the highest level of education that they had completed. It was found that most had completed a graduate degree (36.5%) and the least (5.7%) had only completed diploma as outlined in Table 22.

Table 22: Respondent Education Distribution

	Highest level of Education completed	Frequency N=192	Percentage
1.	Primary School	0	0%
2.	High School	26	13.5%
3.	Diploma	11	5.7%
4.	Undergraduate Degree	59	30.7%
5.	Graduate Degree	70	36.5%
6.	Doctoral Degree	26	13.5%

4.4.4 Role at the University

Respondents were also asked to indicate their role at the university. Majority were staff members at the university (38.5%) but following very closely were students (38%); while the least were faculty (23.4%) as outlined in Table 23.

Table 23: Respondent Role Distribution

	Role at the University	Frequency N=192	Percentage
1.	Student	73	38.0%
2.	Faculty/Lecturer	45	23.4%
3.	Staff	74	38.5%

4.4.5 Years on the Internet

Respondents were asked to recall the year bracket they first used the internet. Majority indicated between 2001 -2005 (31.3%) which represented over 12 to 16 years on the internet. The least number had used the internet before 1991 (3.6%) representing over 27 years on the internet. This is outlined in Table 24.

Table 24: Respondent Year first used the Internet Distribution

	Year first used Internet	Frequency N=192	Percentage
1.	Before 1991	7	3.6%
2.	1991-1995	21	10.9%
3.	1996-2000	55	28.6%
4.	2001-2005	60	31.3%
5.	2006-2010	41	21.4%
6.	after 2010	8	4.2%

4.4.6 Hours on the Internet

Respondents were also asked to indicate how many hours they spent on the internet in a day. Majority spent 5-10 hours in a day (45.8%) while the least spent 21-24 hours in a day (1%) as indicated in Table 25.

Table 25: Respondent Hours on the Internet in a Day Distribution

	Hours spent on the Internet in a day	Frequency N=192	Percentage
1.	Less than 5	48	25%
2.	5-10	88	45.8%
3.	11-15	35	18.2%
4.	16-20	19	9.9%
5.	21-24	2	1%

4.4.7 Computer Skill

Respondents were asked to rate their computer skill on a 5-point Likert scale where 1 is low (meaning that they have little or no skill; requiring a lot of assistance to perform tasks on a computer); 2 is basic (meaning that they can navigate a computer and perform simple tasks such as prepare documents, print and respond to emails); 3 is intermediate (meaning that in addition to basic tasks they can prepare and analyze data on spreadsheets, make presentations with little or no assistance); 4 is advanced (meaning that in addition to intermediate tasks they can change configuration settings, customize applications, backup and manage personal data) and 5 is Expert (meaning that in addition to advanced tasks they can write applications, audit and secure computer systems, troubleshoot and solve computer problems).

Majority rated themselves to have advanced computer skills (40.6%) while the least rated themselves to have basic computer skills (5.7%) and no one rated themselves to have low-level computer skills as outlined in Table 26.

Table 26: Respondent Computer Skills Distribution

	Computer Skill	Frequency N=192	Percentage
1.	Low	0	0%
2.	Basic	11	5.7%
3.	Intermediate	67	34.9%
4.	Advanced	78	40.6%
5.	Expert	36	18.8%

4.4.8 Email Load

Respondents were also asked to indicate how many emails they received in their official email account in a day. Majority indicated less than 10 (43.8%) while the least indicated they received more than 50 emails in a day (3.1%) as outlined in Table 27.

Table 27: Respondent Email Load Distribution

	Emails received in official email in a day	Frequency N=192	Percentage
1.	Less than 10	84	43.8%
2.	11-20	63	32.8%
3.	21-30	21	10.9%
4.	31-40	11	5.7%
5.	41-50	7	3.6%
6.	More than 50	6	3.1%

4.4.9 Email Responsiveness

Respondents were asked to rate their email responsiveness based on two 5-point Likert measures ranging from strongly disagree (1) to strongly agree (5). The first measure was whether they read all the emails they receive in their official email account. The second measure was whether they respond to all emails they need to respond to.

Majority of the respondents strongly agreed (43.2%) to the first measure indicating they read all emails they received in their official email account. Conversely the least number of respondents strongly disagreed (4.7%) that they read all emails received in their official email account. The statistics on the email reading are represented in Table 28.

Table 28: Respondent Email Reading Distribution

	I read all emails I receive in my official email account	Frequency N=192	Percentage
1.	Strongly Disagree	9	4.7%
2.	Disagree	24	12.5%
3.	Neutral	28	14.6%
4.	Agree	48	25%
5.	Strongly Agree	83	43.2%

With regards to the second measure, majority of the respondents indicated that they strongly agreed with the statement (30.7%) indicating that they respond to all emails they need to; while the least strongly disagreed with the statement (8.3%) as outlined in Table 29.

Table 29: Respondent Email Response Distribution

I respond to all emails I need to in my official email account	Frequency N=192	Percentage
1. Strongly Disagree	16	8.3%
2. Disagree	28	14.6%
3. Neutral	33	17.2%
4. Agree	56	29.2%
5. Strongly Agree	59	30.7%

4.4.10 Online Services Usage

The respondents were also asked to rate their usage of four online services (email, social media, online shopping and online banking) on a 5-point Likert scale where 1 is rated as not at all, 2 is little extent, 3 is some extent, 4 is great extent, 5 is very great extent. Regarding email usage, majority indicated they used it to a very great extent (79.7%) while the least indicated to a little extent (1%) while no one indicated 'not at all' (0%) as indicated in Table 30.

Table 30: Respondent Email Usage Distribution

Extent they use Email	Frequency N=192	Percentage
1. Not at all	0	0%
2. Little Extent	2	1%
3. Some Extent	11	5.7%
4. Great Extent	26	13.5%
5. Very Great Extent	153	79.7%

Regarding social media usage, majority indicated they used it to a very great extent (34.9%), closely followed by those who used it to a great extent (34.4%); while the least indicated that they did not use it at all (3.1%) as outlined in Table 31.

Table 31: Respondent Social Media Usage Distribution

Extent they use Social Media	Frequency N=192	Percentage
1. Not at all	6	3.1%
2. Little Extent	18	9.4%
3. Some Extent	35	18.2%
4. Great Extent	66	34.4%
5. Very Great Extent	67	34.9%

Regarding online shopping usage, majority indicated that they used it to a little extent (27.1%), closely followed by those who did not use it at all (26.6%) and to some extent (26.6%); while the least indicated they used it to a very great extent (5.2%) as outlined in Table 32.

Table 32: Respondent Online Shopping Usage Distribution

Extent they use Online Shopping	Frequency N=192	Percentage
1. Not at all	51	26.6%
2. Little Extent	52	27.1%
3. Some Extent	51	26.6%
4. Great Extent	28	14.6%
5. Very Great Extent	10	5.2%

Regarding online banking usage, majority indicated that they did not use it at all (27.1%); while the least indicated they used it to some extent (14.6%) as outlined in Table 33.

Table 33: Respondent Online Banking Usage Distribution

Extent they use Online Banking	Frequency N=192	Percentage
1. Not at all	52	27.1%
2. Little Extent	39	20.3%
3. Some Extent	28	14.6%
4. Great Extent	37	19.3%
5. Very Great Extent	36	18.8%

4.4.11 Prior Victimization

Respondents were also asked to indicate if they had been victims of five online threats in the past. Responses were either, yes, no or I don't know. The first online threat was scams (for example, receiving an email that convinces them to reveal personal details or send money to a falsified recipient). Majority of the respondents indicated that they had been prior victims of scams in the past (82.8%); while the least did not know whether they had been victims (2.1%) as outlined in Table 34.

Table 34: Respondent Prior Victim of Scams Distribution

Prior Victim of Scams	Frequency N=192	Percentage
1. Yes	159	82.8%
2. No	29	15.1%
3. I Don't Know	4	2.1%

The second online threat was online account hijacking (for example, someone taking over their online account and sending messages pretending to be them). Majority of the respondents indicated that they had not been prior victims of online account hijacking (69.8%); while the least did not know whether they had been victims (5.2%) as outlined in Table 35.

Table 35: Respondent Prior Victim of Online Account Hijacking Distribution

	Prior Victimization to Online Account Hijacking	Frequency N=192	Percentage
1.	Yes	48	25%
2.	No	134	69.8%
3.	I Don't Know	10	5.2%

The third online threat was identity theft (for example, someone opens an account taking up another's persona online). Majority of the respondents indicated that they had not been prior victims of identity theft (76.6%); while the least did not know whether they had been victims (10.4%) as outlined in Table 36.

Table 36: Respondent Prior Victim of Identity Theft Distribution

	Prior Victim of Identity Theft	Frequency N=192	Percentage
1.	Yes	25	13%
2.	No	147	76.6%
3.	I Don't Know	20	10.4%

The fourth online threat was credit/debit card fraud (for example, someone uses their credit/debit card to make payments without their knowledge or consent). Majority of the respondents indicated that they had not been prior victims of credit/debit card fraud (89.1%); while the least did not know whether they had been victims (2.6%) as outlined in Table 37.

Table 37: Respondent Prior Victim of Credit/Debit Card Fraud Distribution

	Prior Victim of Credit/Debit Card Fraud	Frequency N=192	Percentage
1.	Yes	16	8.3%
2.	No	171	89.1%
3.	I Don't Know	5	2.6%

The fifth online threat was malicious software infection (for example, their computer gets infected by viruses thus degrading its performance and compromising the confidentiality, integrity and availability data). Majority of the respondents indicated that they had been prior victims of malicious software infections (72.9%); while the least (4.2%) did not know if they had been victims as outlined in Table 38.

Table 38: Respondent Prior Victim of Malicious Software Infection Distribution

Prior Victim of Malicious Software Infection	Frequency N=192	Percentage
1. Yes	140	72.9%
2. No	44	22.9%
3. I Don't Know	8	4.2%

4.4.12 Risk Propensity

Respondents were also asked to rate their risk propensity using three measurement questions rated on 5-point Likert scales. Majority of the respondents rated their risk propensity as neutral meaning they neither agreed nor disagreed regarding the statements on their risk propensity. The respondent risk propensity distribution across the three measurement items is outlined in Table 39.

Table 39: Respondent Risk Propensity Distribution

Measurement Item	(1) Strongly Disagree	(2) Disagree	(3) Neutral	(4) Agree	(5) Strongly Agree
1. I like taking risks	4.7%	13.0%	34.4%	33.3%	14.6%
2. People say I am a risk taker	7.3%	22.9%	31.3%	29.2%	9.4%
3. I take risks that could threaten my safety	25.0%	25.0%	29.7%	13.0%	7.3%

4.5 Exploratory Factor Analysis

Factor analysis is a crucial component of Structural Equation Modeling. Brown (2006) points out that the focus of factor analysis is to reduce a large number of observed indicators into the smaller set of underlying latent constructs by examining the correlations between the observed indicators. It can then be concluded that the observed indicators are basically measuring the same thing but from different dimensions.

Two types of factor analysis should be undertaken: Exploratory Factor Analysis and Confirmatory Factor Analysis. Hassan & Abu Bakar (2009) provide various criteria to show the similarities and differences between the two. Exploratory Factor Analysis does not begin with a pre-determined number of factors but rather lets the dataset inform the factor structure. It is therefore useful for exploratory theory building and testing of new or modified measurement scales. On the other hand, Confirmatory Factor Analysis starts with a predefined indicator-factor structure and then checks whether the dataset conforms to it. The predefined structure is derived from and grounded in some underlying theory.

Hair Jr., Matthews, Matthews, & Sarstedt, (2017) point out that Structural Equation Modeling should include a combination of both exploratory factor analysis and confirmatory factor analysis. In fact, Gaskin (2012, 2019) and Marsh, Morin, Parker, & Kaur (2014) advocate for a combination of both approaches in the development of quantitative models in structural equation modeling.

IBM SPSS Statistics version 23 was used to perform the Exploratory Factor Analysis. The KMO and Bartlett’s test of sphericity and the reproduced correlation matrix were derived in the factor analysis. Factor extraction was done using the Maximum Likelihood method because it maximizes the difference between factors while also providing many helpful indices for assessing the resulting factor structure; particularly indices that can evaluate the goodness-of-fit. No predefined number of factors was defined for extraction; the extraction was based on Eigenvalues greater than 1. Promax rotation method was selected because it is the widely used orthogonal rotation method (Kline, 2013). The Kappa value was set to 4. In addition, small loadings less than ±0.3 were suppressed with the objective of obtaining a clean pattern matrix. According to Hair et al. (2009) factor loadings of ±0.3 to ±0.4 are minimally acceptable although values greater than ±0.5 are more desirable.

4.5.1 Initial Factor Matrix

When all indicators were included in the EFA, 23 factors were extracted explaining a 68.34% variance. In addition, the Kaiser-Meyer-Olkin (KMO) Measure of Sampling Adequacy was 0.739 (which was above the 0.7 threshold for adequacy) and the Bartlett's Test of Sphericity was significant.

4.5.2 Cleaning the Factor Pattern Matrix

There were some indicators that had less than ±0.3 loadings and there were also noticeable cross-loadings. In order to obtain a better quality factor matrix these indicators with poor loadings and cross-loadings were dropped. This meant that the following ten indicators referenced in the code book in Appendix G: QSR_OB1, POL_1, POL_2, POL_3, EM_1, EM_2, EM_3, SE_1, SE_2 and SE_3 had to be dropped. Effectively, this also meant that three factors (Self-efficacy, Policies and Emotions) were dropped from the model.

4.5.3 Total Variance Explained

The resulting factor matrix had 21 factors explaining 68.78% of the total variance as illustrated in Table 40.

Table 40: Final EFA - Total Variance Explained

Factor	**Initial Eigenvalues**			**Extraction Sums of Squared Loadings**			**Rotation Sums of Squared Loadings**[a]
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total
1	11.306	13.622	13.622	9.876	11.899	11.899	6.381
2	9.517	11.467	25.089	8.736	10.525	22.424	7.790
3	6.289	7.577	32.666	4.912	5.918	28.342	5.188
4	4.077	4.912	37.578	5.122	6.171	34.513	6.358
5	3.551	4.279	41.856	3.039	3.661	38.174	5.569
6	3.172	3.821	45.678	1.897	2.285	40.459	4.312
7	2.852	3.436	49.114	2.354	2.837	43.296	3.228
8	2.686	3.236	52.349	1.942	2.340	45.636	5.571
9	2.341	2.820	55.169	1.901	2.291	47.927	4.312
10	2.204	2.655	57.824	1.772	2.135	50.062	6.553
11	1.995	2.403	60.227	2.539	3.059	53.121	3.887
12	1.808	2.178	62.406	1.406	1.694	54.815	6.164
13	1.722	2.075	64.480	1.792	2.159	56.974	5.227
14	1.532	1.846	66.327	1.367	1.647	58.621	4.164
15	1.417	1.707	68.033	1.689	2.034	60.656	4.946
16	1.382	1.665	69.699	1.144	1.378	62.034	2.699
17	1.322	1.593	71.292	1.338	1.612	63.646	3.541
18	1.305	1.572	72.864	1.189	1.432	65.078	2.920
19	1.105	1.331	74.195	1.085	1.307	66.385	4.446
20	1.061	1.279	75.473	1.021	1.231	67.616	3.736
21	1.013	1.221	76.694	.968	1.166	68.782	7.553
22	.923	1.113	77.807				
23	.887	1.068	78.875				
24	.845	1.018	79.893				

4.5.4 Kaiser-Meyer-Olkin (KMO) and Bartlett's Test

The Kaiser-Meyer-Olkin (KMO) Measure of Sampling Adequacy (MSA) was 0.761; which was above the 0.7 threshold for meritorious adequacy as prescribed by Hair et al. (2009). In addition, the Bartlett's Test of Sphericity was statistically significant as required. These statistics are as shown in Table 41.

Table 41: EFA – KMO and Bartlett's Test

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.761
Bartlett's Test of Sphericity	Approx. Chi-Square	12458.323
	df	3403
	Sig.	.000

4.5.5 Goodness-of-fit test

The Goodness-of-fit test was significant and the Chi-square/df ratio (χ^2/df) was 1.197 which is within the desired threshold of between 1 and 3 as prescribed by Hair et al. (2009). This is captured in Table 42.

Table 42: EFA – Goodness-of-Fit Test

Chi-Square	2238.632
df	1870
Chi-square/df ratio	1.197

4.5.6 Residuals

The SPSS statistical software computed residuals between observed and reproduced correlations. The results obtained showed that there were 4% non-redundant residuals with values greater than ±0.05 as shown in Table 43. This is within the recommended threshold where non-redundant residuals should be less than 5%.

Table 43: EFA – Goodness-of-Fit Test

Extraction Method: Maximum Likelihood.
a. Reproduced communalities
b. Residuals are computed between observed and reproduced correlations. There are 141 (4.0%) nonredundant residuals with absolute values greater than 0.05.

4.5.7 Factor Pattern Matrix

The resulting pattern matrix extracted 21 factors from 83 indicators as shown in Table 44. Factor 1 represents a set of Determinants of Trust that require technical knowledge to examine (for example, web address and hyperlink evaluation, website encryption or padlock icon, website certificate, domain registration information and use of security tools). Factor 1 is therefore represented as DT_HIGH in subsequent analysis. Factor 2 matches the seven Persuasive Cues as earlier identified in the theoretical model and is therefore represented as PC. Factor 3 represents Perceived Severity as earlier identified in the theoretical model and is represented as PS. Factor 4 represents a subset of the Determinants of Trust that are poor trust indicators (for example, consistency in logo, colours, look and feel, grammar and spelling, personalized greetings, content reasonableness, expected context and email address of sender). Factor 4 is therefore represented as DT_LOW. Factor 5 represents the outcome variable consisting of both directly observed measures that start with the initial DOB and the questionnaire self-reported measures that start with the prefix QSR. Factor 6 represents Technology Controls from the theoretical model and is represented as TC. Factor 7 represents a combination of indicators from the Pressure and Distraction thus revealing an underlying common factor that was previously identified in the theoretical model as the Ability to Process and is therefore labeled as AP. Factor 8 represents the Elaboration factor and is labeled as ELAB. Factor 9 represents Response Efficacy factor and is labeled as RE. Factor 10 represents the Detection Cues factor and is labeled as DC. Factor 11 represents the threat domain knowledge measured by a Knowledge Quiz and is labeled as KQ. Factor 12 represents the Involvement factor and is labeled INV. Factor 13 represents the Quality of Argument factor and is labeled as QA. Factor 14 represents Perceived Vulnerability factor and is labeled PVUL. Factor 15 represents Security Education Training and Awareness and is labeled SETA. Factor 16 represents the Responsible factor and is labeled RES. Factor 17 represents the Response Cost factor and is labeled RC. Factor 18 represents the Risk Propensity factor and is labeled RP. Factor 19 represents the Perceived Benefit factor and is labeled PB. Factor 20 represents Threat Avoidance factor and is labeled TAV. Factor 21 represents Threat Detection factor and is labeled TD.

The resulting Exploratory Factor Analysis factor pattern matrix was closely matched to the multi-dimensional theoretical model. Majority of the theoretical constructs were found in the resulting pattern matrix. Out of 23 anticipated theoretical model constructs, 20 were extracted and so was 1 factor that would be used as a demographic variable. This indicates that the measurement items associated with the observed indicators were good measures. This provides confidence in reusing the measurement items and scales in future studies. The resulting factor pattern matrix is shown in Table 44.

Table 44: Final EFA – Pattern Matrix

	Factor																				
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
DT11	.952																				
DT13	.816																				
DT10	.798																				
DT12	.750																				
DT9	.678																				
DT8	.398																				
PC4		.925																			
PC5		.902																			
PC7		.787																			
PC6		.700																			
PC2		.597																			
PC3		.560																			
PC1		.530																			
PS4			.889																		
PS3			.830																		
PS2			.821																		
PS1			.759																		
PS6			.551																		
PS5			.550																		
DT2				.816																	
DT3				.747																	
DT1				.739																	
DT4				.723																	
DT5				.589																	
DT6				.559																	
DT7				.326																	
DOB_OB3					.791																
DOB_OB2					.735																
QSR_OB2					.722																
QSR_OB3					.661																
TC2						.980															
TC1						.903															
TC3						.867															
PRES1							.730														
DIST3							.672														
PRES3							.666														
DIST2							.661														
PRES2							.655														
DIST1							.641														
ELAB2								.899													
ELAB1								.832													
ELAB3								.792													
RE2									.824												
RE3									.738												
RE1									.731												
DC2										.887											
DC1										.870											
DC3										.778											
KQ1											.727										
KQ5											.695										
KQ4											.557										
KQ6											.526										
KQ2											.449										
KQ3											.447										
INV1												.907									
INV2												.863									
INV3												.850									
QA3													.934								
QA2													.824								
QA1													.638								
PVUL3														.913							
PVUL2														.877							
PVUL1														.722							
SETA3															.946						
SETA1															.773						

	Factor																				
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
SETA2															.740						
RES3																.835					
RES2																.796					
RES1																.649					
RC2																	.824				
RC1																	.816				
RC3																	.744				
RP2																		.879			
RP1																		.862			
RP3																		.473			
PB2																			.913		
PB3																			.860		
PB1																			.428		
TAV1																				.853	
TAV2																				.821	
TD1																					.834
TD2																					.817
TD3																					.732

Extraction Method: Maximum Likelihood.
Rotation Method: Promax with Kaiser Normalization.

a. Rotation converged in 9 iterations.

4.6 Exploratory Cluster Analysis

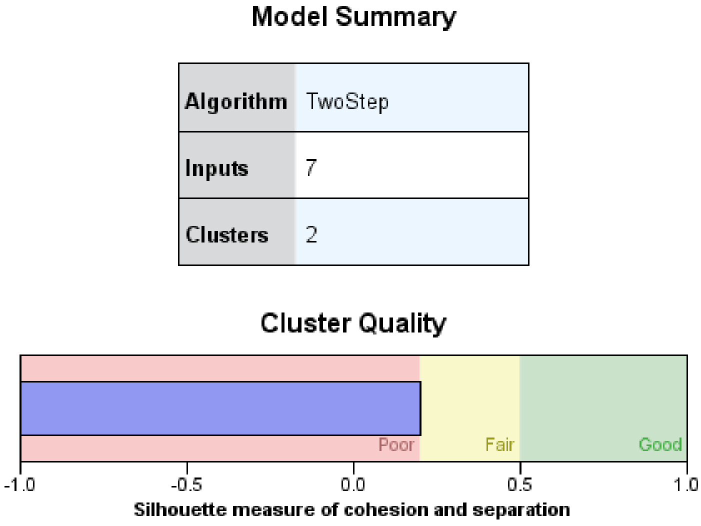
Cluster analysis is a type of exploratory analysis that seeks to organize an empirical dataset into meaningful groups based on a set of variables that best capture the key features in the dataset. Clusters are groups of observations in the dataset that are similar to each other (homogenous) and different from the others based on some characteristics. The objective is therefore to maximize the homogeneity within the cluster while maximizing the heterogeneity among different clusters. Cluster analysis is different from factor analysis in that it is concerned with grouping objects (or cases) into clusters and not grouping variables into factors (Hair et al., 2009; Norusis, 2012).

Cluster analysis does not involve assessment of significance values or testing of hypothesis. It is mainly for demographic profiling and provision of descriptive statistics that can shed more light on the research results. Hair et al. (2009) points out that cluster analysis is very important in examining how different groups of people (demographically profiled and identifiable by certain similarities or differences) differ in the analysis and interpretation of results.

IBM SPSS Statistics version 23 was used to perform the cluster analysis. Three methods were available for use: hierarchical, k-means and two-step. The two-step method was selected because no presumptions on the variables for clustering was made and neither was a fixed number of clusters to extract determined. The software was required to derive clusters from the information in the empirical data.

The first round of cluster analysis used seven demographic variables for clustering (age, education, role, years on the internet, hours on the internet, computer skill and email load). The variables gender, email responsiveness, online services usage, prior victimization and risk propensity were not used because the initial descriptive analysis in Section 5.5 did not reveal much variability in the dataset. The clustering settings used log-likelihood distance measure, Schwarz's Bayesian Criterion (BIC) and the clusters were to be determined automatically with a maximum of 15 clusters allowed.

The initial cluster model had poor cluster quality due to four variables that had less than 0.5 cluster importance as illustrated in Figures 20 and 21. To improve cluster quality the four variables that had least importance were dropped namely: Email load, Year first used the internet, Hours spent on the internet in a day and Computer skills.



Model Summary

Algorithm	TwoStep
Inputs	7
Clusters	2

Figure 20: Initial Cluster Analysis Model Summary

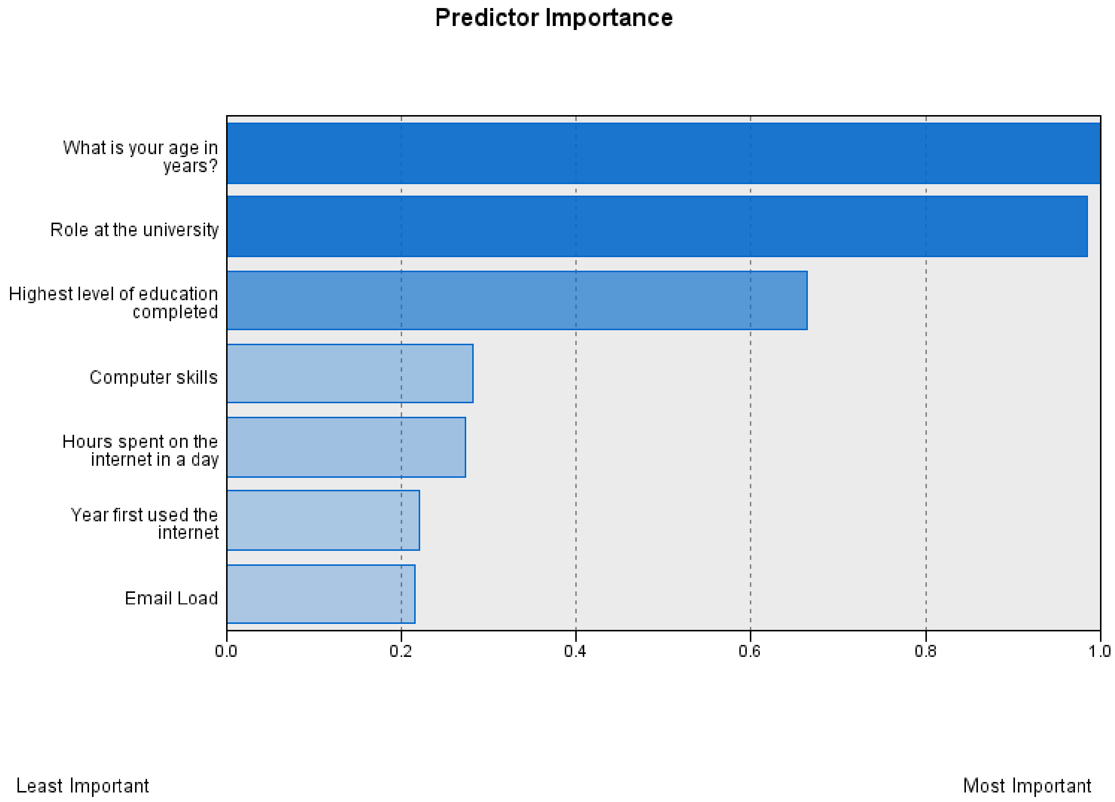
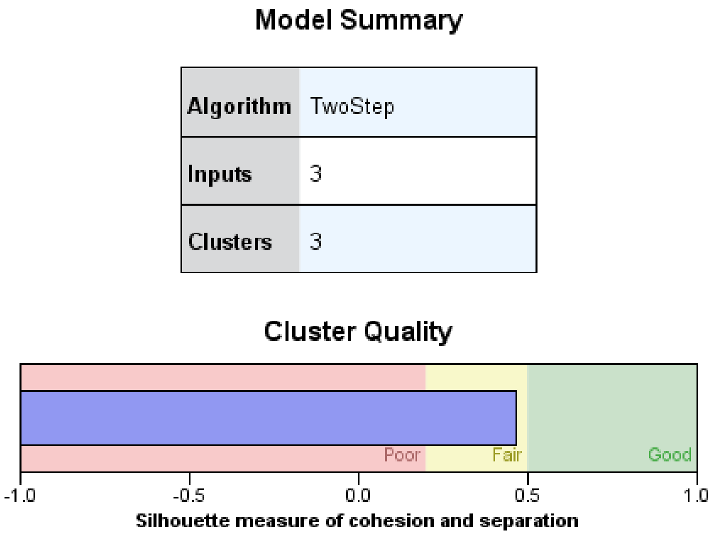


Figure 21: Initial Cluster Predictor Importance

The resulting cluster model, after dropping the four variables of least importance, had much better quality as shown by the Silhouette measure of cohesion and separation in Figure 22.

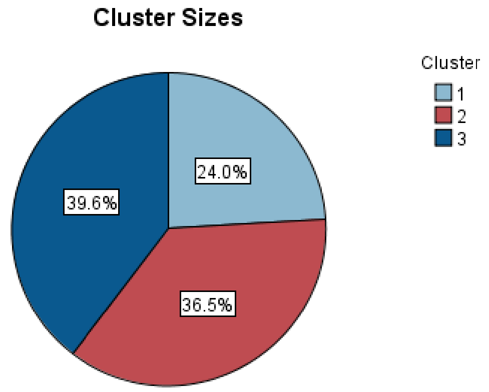


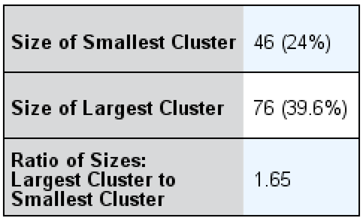
Model Summary

Algorithm	TwoStep
Inputs	3
Clusters	3

Figure 22: Final Cluster Model Quality

The distribution of cases between clusters was also fairly good as indicated by the resulting cluster sizes in Figure 23.

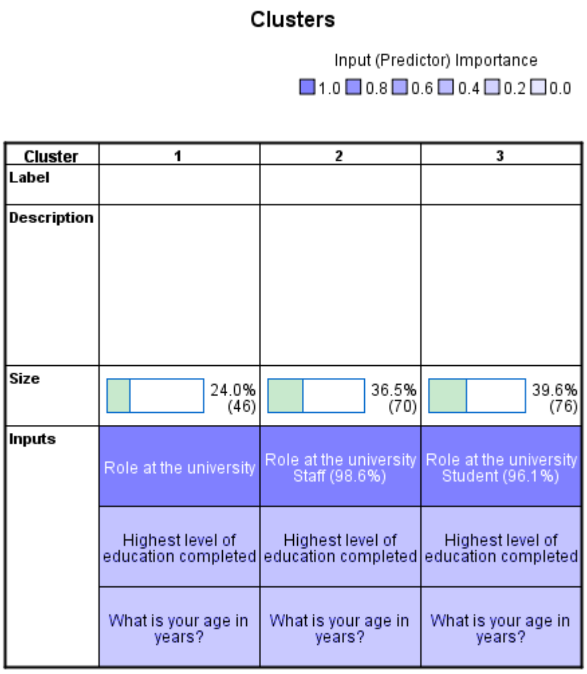




Size of Smallest Cluster	46 (24%)
Size of Largest Cluster	76 (39.6%)
Ratio of Sizes: Largest Cluster to Smallest Cluster	1.65

Figure 23: Final Cluster Model Sizes

Three clusters resulted from the cluster analysis with the role at the university being the most important distinguishing cluster predictor and highest level of education completed being next important and finally followed by age in years. Cluster 1 consisted of faculty/lecturers whose highest level of education was the doctorate degree and their age ranged from 36-45 years. Cluster 2 was made up of staff members whose highest level of education was the graduate degree and their age range was 26-35 years. Finally, cluster 3 was composed of students whose highest level of education was undergraduate degree and whose age ranged from 18-25 years. In terms of cluster sizes, cluster 3 consisting of students was the largest and cluster 1 consisting of faculty/lecturers was the smallest. These details are summarized in Figure 24 on predictor importance and Figure 25 on cluster comparison.



Clusters

Input (Predictor) Importance

□1.0 □0.8 □0.6 □0.4 □0.2 □0.0

Cluster	1	2	3
Label			
Description			
Size	24.0% (46)	36.5% (70)	39.6% (76)
Inputs	Role at the university	Role at the university Staff (98.6%)	Role at the university Student (96.1%)
	Highest level of education completed	Highest level of education completed	Highest level of education completed
	What is your age in years?	What is your age in years?	What is your age in years?

Figure 24: Final Clusters with Predictor Importance

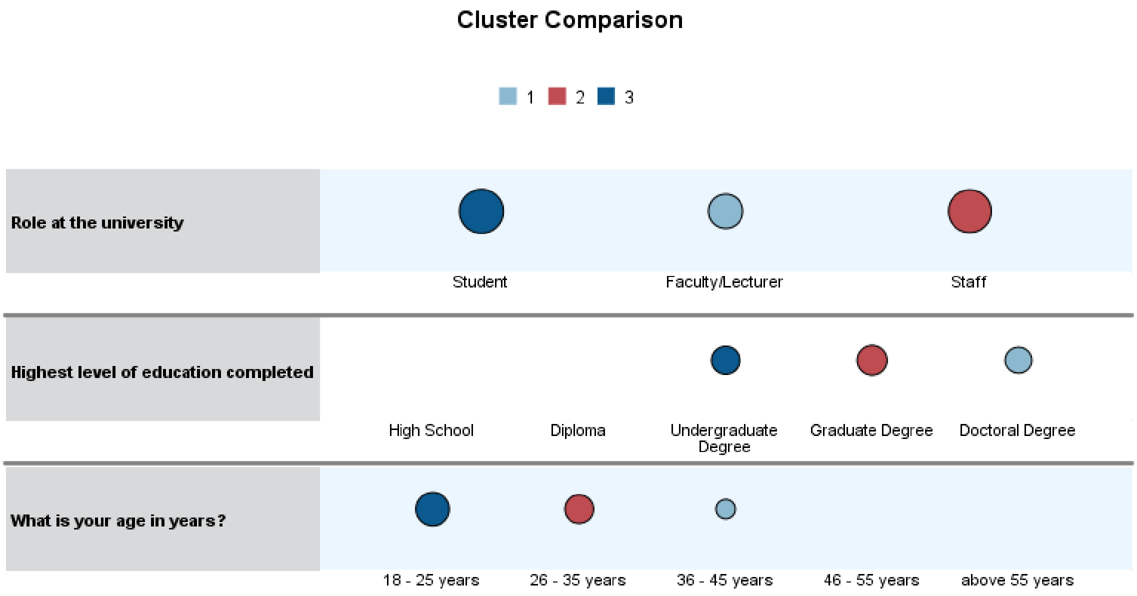


Figure 25: Final Cluster Comparisons

With this final resulting cluster model, a cluster membership variable was created and added to the dataset. This cluster membership variable is useful in subsequent analysis since it is used to group cases into specific demographic groups. This helps evaluate group differences in the inferential statistical analysis.

4.7 Confirmatory Factor Analysis

The Confirmatory Factor Analysis (CFA) is done using IBM AMOS software. Unlike the Exploratory Factor Analysis (EFA), the CFA starts with a predefined indicator-factor structure drawn from theory. The multi-dimensional theoretical model therefore provides the blueprint for the CFA. For that reason, the first step in the CFA is to draw the measurement model based on the theoretical depiction of the multi-dimensional model. However the EFA provides empirical-based input that needs to inform the CFA process. Therefore, the EFA results will also be taken into consideration in the CFA.

4.7.1 CFA Diagram

The Confirmatory Factor Analysis diagram was constructed on IBM SPSS AMOS Graphics, as illustrated in Figure 26, based on the theoretical foundation provided by the multi-dimensional model and also using the empirical results provided by the Exploratory Factor Analysis.

It should be noted that the Exploratory Factor Analysis prompted ten indicators to be dropped from the model. These were: QSR_OB1, POL_1, POL_2, POL_2, EM_1, EM_2, EM_3, SE_1, SE_2 and SE_3. Effectively, this meant that three variables; Self-Efficacy (SE), Policies (POL) and Emotions (EM) were dropped from the theoretical model. In addition, an underlying latent variable called 'Ability to Process' was constructed from six indicators that were initially hypothesized to represent two different variables; Distraction (DIST) and Pressure (PRES). Furthermore, one additional factor was added to the theorized model by splitting the 'Determinants of Trust' factor into two dimensions; Low Determinants of Trust (DT_LOW) and High Determinants of Trust (DT_HIGH).

The measurement model path diagram constructed in Amos Graphics for Confirmatory Factor Analysis is as illustrated in Figure 26 and following best practice guidelines (Awang, 2012, 2015; Kline, 2013). The path diagram consists of the 21 latent factors identified in the Exploratory Factor Analysis with their associated indicator measures. The indicator measures are depicted as rectangles with reflective arrows coming from the latent constructs, which are depicted as ovals, to the indicator measures. Error terms, depicted as small circles on the right-hand side, are linked to each rectangle-shaped indicator. In addition, double-sided covariance arrows are used to link the oval-shaped latent constructs with each other.

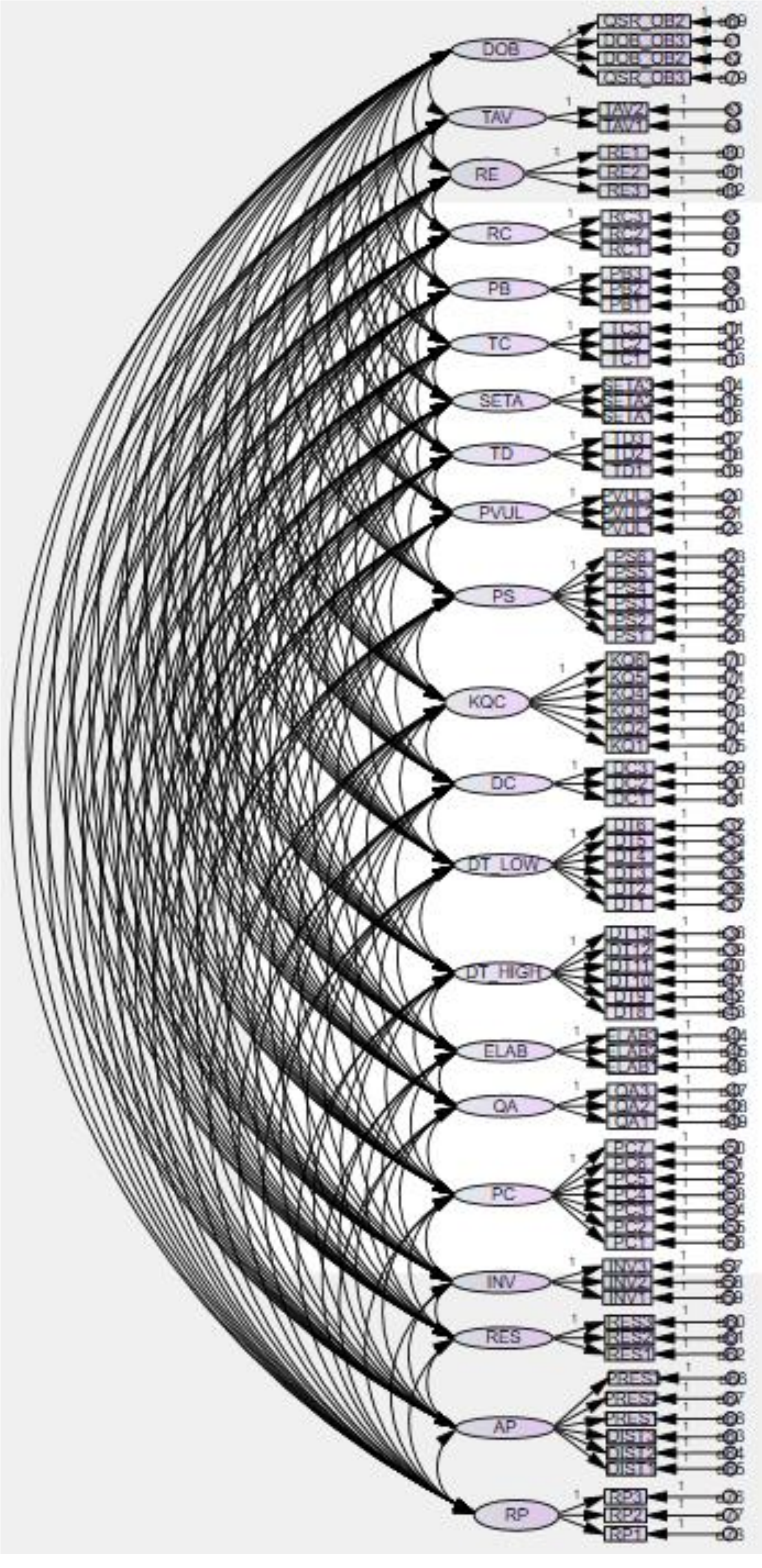


Figure 26: Amos Graphics CFA Diagram

4.7.2 Unidimensionality

Factor unidimensionality was assessed by looking at the factor loadings of each indicator. Awang (2012) explains that the factor loadings should be above 0.5 and they should be positive. If any factor loading is below the 0.5 threshold, the associated indicator should be deleted and the model re-run until all factor loadings are satisfactory in order to obtain unidimensionality. Table 45 shows the 21 factors and the factor loadings associated with their respective indicators.

It was found that KQ3 (0.478), KQ6 (0.468), PC2 (0.434) and RP3 (0.481) had unsatisfactory loadings. Deletion began by removing the indicator with the lowest loading PC2 followed by KQ6, KQ3 and finally RP3. The model was re-run after removing an indicator and loadings were reviewed before deleting the next indicator. After deleting RP3 another indicator RP1 fell below the threshold because it had a loading of 0.424. This would have required that the Risk Propensity (RP) factor be dropped. However, it was decided to retain the RP indicators and factor and further observations would be made regarding any reliability and validity concerns. Therefore, an additional 3 indicators were dropped out of the initial 83 indicators accounting for 3.66% deletion. Awang (2012) warns that deletion through this process should not exceed 20% of total indicator items. The deletion was within this threshold and therefore there were no concerns raised.

Table 45: CFA – Factor Unidimensionality

	Factor	Indicators	Factor Loading
1.	UIT Outcome Behavioural (DV_OB)	QSR_OB2	0.683
		QSR_OB3	0.56
		DOB_OB2	0.926
		DOB_OB3	0.948
2.	Threat Avoidance (TAV)	TAV1	0.929
		TAV2	0.875
3.	Response Efficacy (RE)	RE1	0.861
		RE2	0.866
		RE3	0.782
4.	Response Cost (RC)	RC1	0.815
		RC2	0.746
		RC3	0.847
5.	Perceived Benefit (PB)	PB1	0.745
		PB2	0.948
		PB3	0.909
6.	Technology Controls (TC)	TC1	0.944
		TC2	0.975
		TC3	0.904
7.	Security Education Training and Awareness (SETA)	SETA1	0.812
		SETA2	0.840
		SETA3	0.967

	Factor	Indicators	Factor Loading
8.	Threat Detection (TD)	TD1	0.876
		TD2	0.956
		TD3	0.867
9.	Perceived Vulnerability (PVUL)	PVUL1	0.739
		PVUL2	0.843
		PVUL3	0.933
10.	Perceived Severity (PS)	PS1	0.802
		PS2	0.845
		PS3	0.867
		PS4	0.842
		PS5	0.545
		PS6	0.578
11.	Knowledge Quiz (KQC)	KQ1	0.65
		KQ2	0.534
		KQ3	0.478
		KQ4	0.682
		KQ5	0.609
		KQ6	0.468
12.	Detection Cues (DC)	DC1	0.906
		DC2	0.942
		DC3	0.895
13.	Trust Determinants (DT_LOW)	DT1	0.751
		DT2	0.764
		DT3	0.766
		DT4	0.847
		DT5	0.692
		DT6	0.593
14.	Trust Determinants (DT_HIGH)	DT8	0.513
		DT9	0.785
		DT10	0.892
		DT11	0.91
		DT12	0.734
		DT13	0.816
15.	Elaboration (ELAB)	ELAB1	0.874
		ELAB2	0.96
		ELAB3	0.908
16.	Quality of Argument	QA1	0.825
		QA2	0.862
		QA3	0.903
17.	Persuasive Cues	PC1	0.615
		PC2	0.434
		PC3	0.696
		PC4	0.907
		PC5	0.892
		PC6	0.81
		PC7	0.839
18.	Involvement (INV)	INV1	0.918
		INV2	0.918
		INV3	0.858
19.	Responsible (RES)	RES1	0.616
		RES2	0.803
		RES3	0.866
20.	Ability to Process (AP)	DIST1	0.672
		DIST2	0.617
		DIST3	0.734
		PRES1	0.686
		PRES2	0.645
		PRES3	0.632
21.	Risk Propensity	RP1	0.859
		RP2	0.874
		RP3	0.481

4.7.3 Reliability

Factor reliability was assessed by calculating the Cronbach's Alpha (α) in IBM SPSS Statistics and Composite Reliability (CR) in IBM SPSS AMOS for each factor based on their associated indicators. Hair et al. (2009) point out that the Cronbach's Alpha (α) is the most widely used reliability measure. Fornell & Larcker (1981) and Hair et al. (2009) explain that the Cronbach's Alpha (α) should be greater than 0.70. Nunnally & Bernstein (1994) set the Composite Reliability (CR) threshold to 0.70 as well. The results of the reliability analysis displayed in Table 46 shows that all factors met the reliability criteria.

Table 46: CFA – Factor Reliability

	Factor	Indicators	Factor Loading	Cronbach's Alpha (α)	Composite Reliability (CR)
1.	UIT Outcome Behavioural (DOB)	QSR_OB2 QSR_OB3 DOB_OB2 DOB_OB3	0.683 0.561 0.926 0.948	0.816	0.869
2.	Threat Avoidance (TAV)	TAV1 TAV2	0.929 0.876	0.896	0.898
3.	Response Efficacy (RE)	RE1 RE2 RE3	0.859 0.868 0.783	0.873	0.876
4.	Response Cost (RC)	RC1 RC2 RC3	0.817 0.748 0.843	0.843	0.845
5.	Perceived Benefit (PB)	PB1 PB2 PB3	0.745 0.948 0.909	0.894	0.904
6.	Technology Controls (TC)	TC1 TC2 TC3	0.944 0.975 0.904	0.958	0.959
7.	Security Education Training and Awareness (SETA)	SETA1 SETA2 SETA3	0.812 0.840 0.967	0.902	0.907
8.	Threat Detection (TD)	TD1 TD2 TD3	0.876 0.956 0.867	0.926	0.928
9.	Perceived Vulnerability (PVUL)	PVUL1 PVUL2 PVUL3	0.739 0.844 0.933	0.869	0.879
10.	Perceived Severity (PS)	PS1 PS2 PS3 PS4 PS5 PS6	0.802 0.845 0.867 0.842 0.544 0.578	0.876	0.887
11.	Knowledge Quiz (KQC)	KQ1 KQ2 KQ4 KQ5	0.703 0.580 0.570 0.687	0.725	0.731
12.	Detection Cues (DC)	DC1 DC2 DC3	0.907 0.941 0.895	0.938	0.939

Factor	Indicators	Factor Loading	Cronbach's Alpha (α)	Composite Reliability (CR)
13. Trust Determinants (DT_LOW)	DT1	0.751	0.874	0.878
	DT2	0.765		
	DT3	0.766		
	DT4	0.846		
	DT5	0.692		
	DT6	0.593		
14. Trust Determinants (DT_HIGH)	DT8	0.513	0.901	0.904
	DT9	0.785		
	DT10	0.892		
	DT11	0.910		
	DT12	0.734		
	DT13	0.816		
15. Elaboration (ELAB)	ELAB1	0.874	0.937	0.939
	ELAB2	0.96		
	ELAB3	0.908		
16. Quality of Argument	QA1	0.825	0.896	0.898
	QA2	0.862		
	QA3	0.904		
17. Persuasive Cues	PC1	0.612	0.910	0.912
	PC3	0.693		
	PC4	0.908		
	PC5	0.89		
	PC6	0.812		
	PC7	0.838		
18. Involvement (INV)	INV1	0.918	0.925	0.926
	INV2	0.918		
	INV3	0.858		
19. Responsible (RES)	RES1	0.615	0.790	0.810
	RES2	0.803		
	RES3	0.866		
20. Ability to Process (AP)	DIST1	0.667	0.826	0.827
	DIST2	0.618		
	DIST3	0.720		
	PRES1	0.690		
	PRES2	0.647		
	PRES3	0.651		
21. Risk Propensity	RP1	0.865	0.762	0.795
	RP2	0.868		
	RP3	0.482		

4.7.4 Convergent Validity

Factor convergent validity is assessed using the Average Variance Extracted (AVE) measure calculated in IBM SPSS AMOS for every factor. Awang (2012) prescribes that the AVE should be at least 0.5 for each factor because this means that 50% of the variance in the indicators is accounted for by the factor and not by measurement error. Table 47 shows the first AVE measures for each factor (in the column marked Prior AVE). Two factors did not meet the 0.5 threshold and these are the Ability to Process (AP) and the Knowledge Quiz (KQC) factors. Low factor loadings are known to cause unsatisfactory AVE values. Therefore, to fix the convergent validity concerns, the low loading factors were dropped until satisfactory

AVE values were obtained. This meant that the indicators DIST2, DIST1, PRES3 were deleted from the AP factor and the indicator KQ4 was dropped from the KQC factor in order to fix the convergent validity concerns. The resulting AVE values are displayed in Table 47 in the column marked final AVE. All values are above 0.5 and therefore the factors have demonstrated satisfactory convergent validity.

Table 47: CFA – Convergent Validity

Factor	CR	Prior AVE	Final AVE
DOB	0.869	0.634	0.634
TAV	0.898	0.815	0.815
RE	0.876	0.701	0.701
RC	0.845	0.646	0.646
PB	0.904	0.760	0.760
TC	0.959	0.886	0.886
SETA	0.908	0.767	0.767
TD	0.928	0.811	0.811
PVUL	0.879	0.710	0.709
PS	0.887	0.575	0.574
KQC	0.773	0.407	0.647
DC	0.939	0.836	0.836
DT_LOW	0.878	0.547	0.547
DT_HIGH	0.904	0.618	0.618
ELAB	0.939	0.837	0.837
QA	0.898	0.747	0.747
PC	0.912	0.639	0.639
INV	0.926	0.807	0.807
RES	0.810	0.591	0.591
AP	0.759	0.444	0.512
RP	0.796	0.578	0.579

4.7.5 Discriminant validity

Discriminant validity between factors considers the extent which a factor is distinct and different from all other factors in the model. A factor should not be highly correlated with another factor because if they are they should in fact be part of the same latent factor. In addition, a factor should account for more variance from the indicators associated with it than with indicators associated with other factors. Fornell & Larcker (1981) prescribe the use of the square root of AVE to assess discriminant validity. They explain that the square root of AVE for each factor should be greater than any correlation between any pair of factors in the model.

This is represented by a correlation matrix where the square root of AVE is depicted on the diagonal and the correlations are depicted in the off-diagonal values as illustrated in Table 48. It should be noted that all the validity criteria specified for the model are satisfactorily met.

Table 48: CFA – Discriminant Validity

	CR	AVE	MSV	ASV	AP	DOB	TAV	RC	PB	TC	SETA	TD	PVUL	PS	DC	DT_LOW	DT_HIGH	ELAB	QA	PC	INV	RES	KQC	RP	RE
AP	0.759	0.512	0.051	0.010	**0.716**																				
DOB	0.869	0.634	0.238	0.044	-0.012	**0.796**																			
TAV	0.898	0.815	0.133	0.051	-0.141	-0.026	**0.903**																		
RC	0.845	0.646	0.113	0.029	0.087	-0.115	-0.182	**0.804**																	
PB	0.904	0.760	0.346	0.052	-0.057	0.052	0.290	-0.251	**0.872**																
TC	0.959	0.886	0.347	0.033	0.139	-0.124	0.114	0.149	0.066	**0.941**															
SETA	0.908	0.767	0.347	0.056	0.066	-0.189	0.162	0.121	-0.021	0.589	**0.876**														
TD	0.928	0.811	0.392	0.104	-0.065	-0.488	0.232	0.230	0.098	0.289	0.353	**0.901**													
PVUL	0.879	0.709	0.167	0.038	0.069	-0.040	0.291	-0.197	0.383	0.050	0.056	0.080	**0.842**												
PS	0.887	0.574	0.255	0.041	0.004	-0.150	0.365	-0.336	0.505	-0.015	0.021	0.030	0.279	**0.758**											
DC	0.939	0.836	0.392	0.087	-0.129	-0.250	0.248	0.181	0.087	0.226	0.429	0.626	0.136	0.043	**0.915**										
DT_LOW	0.878	0.547	0.367	0.047	0.011	0.199	0.222	-0.216	0.178	0.062	0.007	-0.141	0.113	0.109	0.044	**0.740**									
DT_HIGH	0.904	0.618	0.185	0.059	-0.094	-0.245	0.241	0.090	0.035	0.206	0.239	0.429	-0.013	-0.015	0.430	0.339	**0.786**								
ELAB	0.939	0.837	0.398	0.072	-0.086	-0.232	0.272	-0.108	0.100	0.063	0.335	0.451	0.222	0.091	0.446	0.150	0.398	**0.915**							
QA	0.898	0.747	0.398	0.059	-0.189	-0.091	0.278	-0.023	0.032	0.091	0.254	0.361	0.143	-0.016	0.370	0.175	0.393	0.631	**0.864**						
PC	0.912	0.639	0.367	0.068	0.004	0.357	0.196	-0.220	0.115	-0.134	-0.195	-0.474	0.199	0.029	-0.286	0.606	0.007	-0.012	0.083	**0.799**					
INV	0.926	0.807	0.335	0.060	0.044	0.423	0.049	-0.243	0.061	-0.084	-0.139	-0.496	0.133	-0.053	-0.309	0.362	-0.245	-0.054	0.060	0.579	**0.898**				
RES	0.810	0.591	0.075	0.024	0.011	-0.089	0.219	0.045	0.103	0.120	0.190	0.149	0.176	0.138	0.223	0.080	0.150	0.273	0.266	0.037	0.072	**0.769**			
KQC	0.773	0.647	0.069	0.015	0.021	-0.112	0.114	-0.013	0.158	0.064	0.126	0.209	0.190	0.210	0.262	0.017	0.060	0.067	0.042	-0.098	-0.124	-0.082	**0.804**		
RP	0.796	0.579	0.086	0.021	-0.070	-0.065	0.139	-0.017	0.147	0.059	0.134	0.158	0.092	0.122	0.293	0.180	0.220	0.207	0.093	0.008	-0.094	0.183	0.041	**0.761**	
RE	0.876	0.701	0.346	0.048	-0.226	-0.040	0.360	-0.148	0.588	-0.005	-0.103	0.097	0.409	0.332	0.065	0.119	0.068	-0.001	-0.129	0.116	0.000	0.110	0.050	0.214	**0.838**

Validity Criteria:

CR stands for Composite Reliability. CR should be greater than 0.7 and should also be greater than the AVE (Nunnally & Bernstein, 1994)
AVE stands for Average Variance Extracted. AVE should be greater than 0.5 (Fornell & Larcker, 1981; Hair et al., 2009)
MSV stands for Maximum Shared Variance.
ASV stands for Average Shared Variance.
MSV should be less than Average Variance Extracted (AVE). ASV should be less than MSV (Hair et al., 2009)
After ASV column is the correlation matrix with the square root of AVE on the diagonal (values in bold)
Square root of AVE values (in bold on the diagonal) should be larger than off-diagonal correlation values (Fornell & Larcker, 1981)

4.7.6 Model Fit

Various model fit indices were examined as prescribed by Hair et al. (2009) in the 3 general groupings of absolute, incremental and parsimony fit measures. The fit measures presented are dependent on the IBM SPSS AMOS output. Generally, one fit measure was provided for each category.

The fit indices were calculated for the measurement model after the adjustments made for unidimensionality, reliability, convergent validity and discriminant validity. The goodness-of-fit values obtained are presented in Table 49 alongside their desired thresholds.

Table 49: CFA – Initial Goodness-of-Fit Indices

Goodness-of-Fit Measure	**Notation**	**Desired Threshold**	**Value Obtained**
Absolute Measures			
Chi-square	χ^2		3688.295
Degrees of freedom	df		2418
Chi-square/df ratio	χ^2/df	< 3	1.525
Standardized Root Mean Residual	SRMR	$< .1$	0.0566
Root Mean Square Error of Approximation	RMSEA	$< .05$	0.052
Incremental Measures			
Comparative Fit Index	CFI	$> .90$	0.881
Parsimonious Fit Measures			
Parsimony Normed Fit Index	PNFI	$> .50$	0.648

The absolute and parsimonious fit indices were satisfactory. However, the Comparative Fit Index chosen to check incremental fit was close to the desired threshold but did not exactly meet the criteria. Therefore, modification indices were calculated and adjustments were made to improve the model fit. Modification indices suggest ways to improve the model. Awang (2012) demonstrates that correlating error terms belonging to the same factor can improve the model fit. This method was used and six pairs of error terms were correlated and this resulted in the desired goodness-of-fit indices as outlined in Table 50. The Chi-square/df ratio, RMSEA, CFI and PNFI values improved thereby confirming a better fitting measurement model.

Table 50: CFA – Final Goodness-of-Fit Indices

Goodness-of-Fit Measure	Notation	Desired Threshold	Value Obtained
Absolute Measures			
Chi-square	χ^2		3421.025
Degrees of freedom	df		2412
Chi-square/df ratio	χ^2/df	< 3	1.418
Root Mean Square Error of Approximation	RMSEA	$< .05$	0.047
Incremental Measures			
Comparative Fit Index	CFI	$> .90$	0.905
Parsimonious Fit Measures			
Parsimony Normed Fit Index	PNFI	$> .50$	0.664

4.8 Structural Model Analysis and Results

The Confirmatory Factor Analysis process provided a measurement model with good factor unidimensionality, reliability, convergent validity, discriminant validity and model fit. The next step in the Structural Equation Modeling process was then to specify and validate the structural model. This involves constructing the structural path diagrams, analyzing model fit and evaluating hypothesis.

4.8.1 Structural Path Model

The validated measurement model was reorganized into a structural path model representing the theorized multi-dimensional model for determining susceptibility to Unintentional insider threats.

A structural path model was defined using the first-order latent factors that emerged from the Exploratory Factor Analysis process. The theorized cause-effect relationships were modeled using regression path lines as theorized in the multi-dimensional model. These regression path lines are depicted as single-headed arrows pointing from one construct to the other in line with the theorized causal relationship. All constructs, including demographic factors, were incorporated into the path model on IBM SPSS Amos Graphics is as shown in Figure 27.

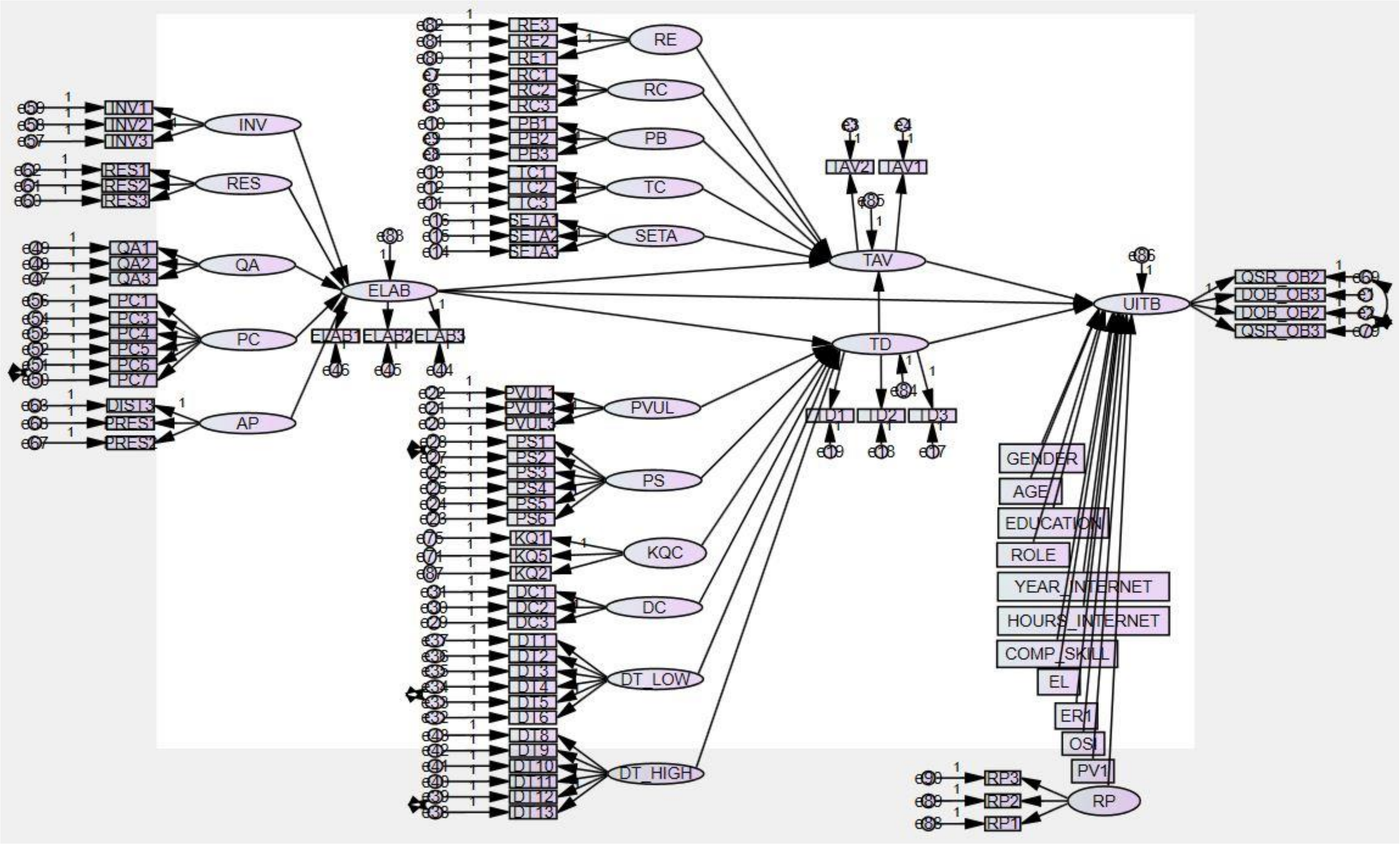


Figure 27: AMOS Graphics Structural Path Model

4.8.2 Structural Model Fit

Similar to the CFA process, the structural path model fit was examined using various indices. Hair et al. (2009) emphasizes that measurement model fit has to be achieved before proceeding with the structural model fit analysis because the fit will not improve when the structural model is specified.

Not all indices met the required thresholds; particularly those relating to incremental fit. This is however not a problem because Hair et al. (2009) point out that only three to four fit indices are needed to judge adequacy of goodness-of-fit. In addition, they point out that complex models with samples less than 250 require less strict criteria when evaluating goodness-of-fit indices. For example, the RMSEA threshold can be relaxed to accept values <.07 and the p-value for Chi-square/df ratio is expected to be significant. The satisfactory indices verifying absolute and parsimonious goodness-of-fit are as outlined in Table 51 alongside the desired thresholds.

Table 51: Structural Model Goodness-of-Fit Indices

Goodness-of-Fit Measure	Notation	Desired Threshold	Value Obtained
Absolute Measures			
Chi-square	χ^2		6092.579
Degrees of freedom	df		3540
Chi-square/df ratio	χ^2/df	< 3	1.721
Root Mean Square Error of Approximation	RMSEA	$< .07$	0.061
Parsimonious Fit Measures			
Parsimony Normed Fit Index	PNFI	$> .50$	0.581
Parsimony Normed Fit Index	PCFI	$> .50$	0.755

The structural model is therefore judged as satisfactory using these four fit indices and with an understanding that the presented model is complex and the sample is less than 250 cases.

4.8.3 Hypotheses Testing

The hypotheses outlined in Section 2.4.2 on the unified multi-dimensional theoretical model for determining susceptibility to unintentional insider threats are outlined and described in Table 52.

Table 52: Proposed Hypotheses

Hypothesis and Description	
H1	Threat Avoidance has a negative and significant effect on the Unintentional Insider Threat Behaviour Outcome
H2a	Response Efficacy has a positive and significant effect on Threat Avoidance
H2b	Self-Efficacy has a positive and significant effect on Threat Avoidance
H2c	Perceived Response Cost has a negative and significant effect on Threat Avoidance
H2d	Perceived Response Benefit has a positive and significant effect on Threat Avoidance
H3a	Policies have a positive and significant effect on Threat Avoidance
H3b	Technology Controls have a positive and significant effect on Threat Avoidance
H3c	Security Education Training and Awareness have a positive and significant effect on Threat Avoidance
H4	Threat Detection has a positive and significant effect on Threat Avoidance
H5	Threat Detection has a negative and significant effect on the Unintentional Insider Threat Behaviour Outcome
H6a	Perceived Vulnerability has a positive and significant effect on Threat Detection
H6b	Perceived Severity has a positive and significant effect on Threat Detection
H7a	Knowledge on Threat Domain has a positive and significant effect on Threat Detection
H7b	Knowledge on Detection Cues has a positive and significant effect on Threat Detection
H7c	Knowledge on Determinants of Trust has a positive and significant effect on Threat Detection
H8	Elaboration has a positive and significant effect on Threat Detection
H9	Elaboration has a positive and significant effect on Threat Avoidance
H10	Elaboration has a negative and significant effect on the Unintentional Insider Threat Behavioural Outcome
H11a	Quality of Argument has a positive and significant effect on Elaboration
H11b	Persuasive Cues have a negative and significant effect on Elaboration
H12a	Involvement has a positive and significant effect on Elaboration
H12b	Responsibility has a positive and significant effect on Elaboration
H13a	Distractions have a negative and significant effect on Elaboration
H13b	Emotions have a negative and significant effect on Elaboration
H13c	Pressure has a negative and significant effect on Elaboration

It is important to point out that three latent variables were dropped from the model during the Exploratory Factor Analysis process. These are the Self-Efficacy, Policies and Emotions latent variables. This shows that the items used to measure these variables were not good enough to capture them satisfactorily. Therefore H2b, H3a and H13b were removed from hypotheses testing. In addition, the Distractions variable and Pressure variable were found to load on one latent factor which was named as Ability to Process. Therefore, the overall H13 relating to Ability to Process was evaluated instead of evaluating the constituent H13a and H13c separately.

Conversely, one variable that was not in the model was added after the Trust Determinants latent variable was split into two dimensions (low trust determinants and high trust determinants. Therefore, H7c was used to test low trust determinants and H7d was used to test high trust determinants.

The hypotheses were tested by examining the standardized path coefficients (β values) to see if they were in the predicted direction. If the path coefficients are greater than 0 they represent positive relationships and conversely if they are less than 0 they represent a negative relationship. In addition, hypotheses testing examined the significance of these hypothesized relationships. Significance was determined by examining the p-values associated with the path coefficients. Significance was tested at $p \leq 0.1$, $p \leq 0.05$ and $p \leq 0.001$ levels and a conclusion indicating that the hypothesis was supported if the p-value was significant.

The path coefficients, direction of relationship and conclusion of the hypotheses testing are outlined in Table 53.

Table 53: SEM Hypothesis Testing Results

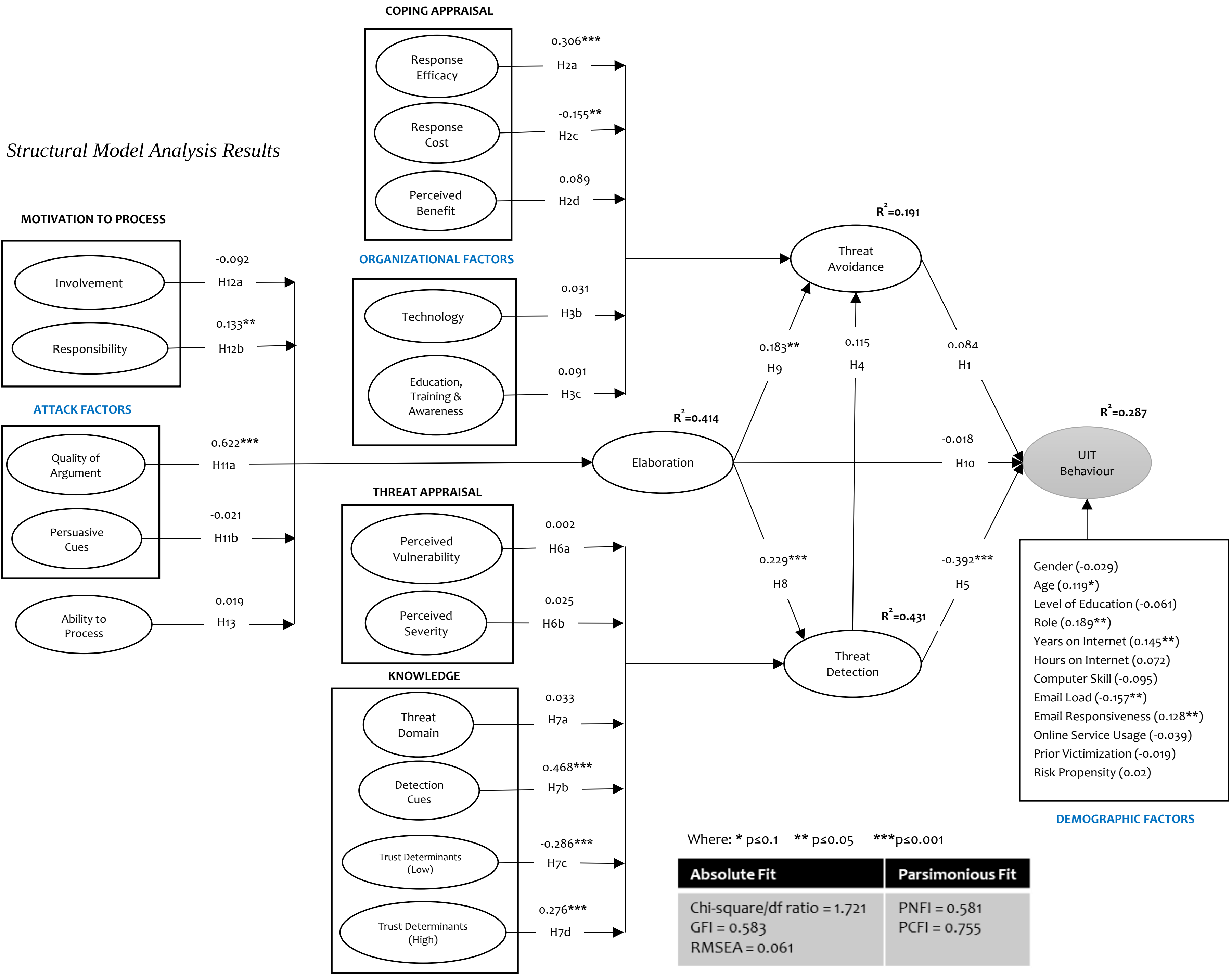
Hypothesis	Path	Comment	β values	S.E.	p-value	Conclusion
H1	TAV -> UITB		0.084	0.033	0.231	Not Supported
H2a	RE -> TAV		0.306	0.114	***	Supported
H2b	SE -> TAV	Dropped at EFA	-	-	-	Removed
H2c	RC -> TAV		-0.155	0.087	0.046	Supported
H2d	PB -> TAV		0.089	0.134	0.225	Not Supported
H3a	POL -> TAV	Dropped at EFA	-	-	-	Removed
H3b	TC -> TAV		0.031	0.058	0.665	Not Supported
H3c	SETA -> TAV		0.091	0.054	0.209	Not Supported
H4	TD -> TAV		0.115	0.054	0.129	Not Supported
H5	TD -> UITB		-0.392	0.024	***	Supported
H6a	PVUL -> TD		0.002	0.077	0.973	Not Supported
H6b	PS -> TD		0.025	0.14	0.685	Not Supported
H7a	KQC -> TD		0.033	0.235	0.63	Not Supported
H7b	DC -> TD		0.468	0.066	***	Supported
H7c	DT_LOW -> TD		-0.286	0.122	***	Supported
H7d	DT_HIGH -> TD	Added at EFA	0.276	0.069	***	Supported
H8	ELAB -> TD		0.229	0.058	***	Supported
H9	ELAB -> TAV		0.183	0.051	0.015	Supported
H10	ELAB -> UITB		-0.018	0.022	0.8	Not Supported
H11a	QA -> ELAB		0.622	0.07	***	Supported
H11b	PC -> ELAB		-0.021	0.072	0.733	Not Supported
H12a	INV -> ELAB		-0.092	0.063	0.14	Not Supported
H12b	RES -> ELAB		0.133	0.102	0.043	Supported
H13	AP -> ELAB	Combined at EFA	0.019	0.114	0.785	Not Supported
H13a	DIST -> ELAB	Dropped at EFA	-	-	-	Removed
H13b	EM -> ELAB	Dropped at EFA	-	-	-	Removed
H13c	PRES -> ELAB	Dropped at EFA	-	-	-	Removed

Results show that a total of 22 hypotheses were tested and 10 of these were supported in the structural model analysis while 12 were not supported by the provided model specification and sample dataset.

Additionally, the coefficient of determination (R^2) values were examined for each endogenous construct to determine the amount of variance that the model was able to explain. It was found that the model was able to explain 41.4% of the Elaboration variance, 43.1% of Threat Detection variance, 19.1% of Threat Avoidance variance and more importantly 28.7% of Unintentional Insider Threat Behaviour variance.

The full results of the structural model analysis showing all model variables, demographic variables, path coefficients, hypotheses testing, R^2 values and goodness-of-fit indices are depicted in Figure 28.

Figure 28: Structural Model Analysis Results



Where: * p≤0.1 ** p≤0.05 ***p≤0.001

Absolute Fit	Parsimonious Fit

Chi-square/df ratio = 1.721
GFI = 0.583
RMSEA = 0.061
	PNFI = 0.581
	PCFI = 0.755

4.9 Chapter Summary

This chapter has outlined the data analysis procedures applied to the collected dataset and has presented the results obtained. The chapter has reported the response rates from both the naturalistic field study and the questionnaire survey used to collect data. It has presented the data screening procedures used to address issues relating to missing values, outliers, common method bias and assumptions of normality.

An overview of the data has been given through descriptive analysis of the demographic variables. In addition, an exploratory cluster analysis has given a meaningful grouping of the cases that characterize the dataset.

Results of the Exploratory Factor Analysis have shown how observed indicators relating to the theoretical model can be summarized using 20 latent variables. This EFA process saw three variables (Self-Efficacy, Policies and Emotions) being dropped from the model due to unsatisfactory factor loadings. In addition, two variables (Distractions and Pressure) proposed in the theoretical model were combined to one latent variable (Ability to Process) based on the factor structure. The EFA also split one proposed variable (Trust Determinants) into two latent factors representing two different dimensions (Low Trust Determinants and High Trust Determinants) of the variable.

Thereafter a Confirmatory Factor Analysis was done and results presented showed a satisfactory measurement model that meets the requirements for unidimensionality, reliability, convergent validity, discriminant validity and goodness-of-fit. The satisfactory measurement model was thereafter translated into a structural model depicting the theorized cause-effect relationships and hypothesis. Validation of the structural model showed adequate goodness-of-fit allowing for hypothesis testing. A total of 22 hypotheses were tested and 10 of these hypotheses were supported by the proposed theoretical model structure. The resulting multi-dimensional model was able to explain 43.1% of the Threat Detection variance, 41.4% of the Elaboration variance, 19.1% of the Threat Avoidance variance and 28.7% of the Unintentional Insider Threat Behaviour variance.

The next chapter dissects these finding and presents a discussion and interpretation of each finding in line with extant Unintentional Insider Threat literature.

CHAPTER 5: DISCUSSIONS

5.1 Introduction

This chapter discusses the findings from this research. It starts by reiterating the objectives set out for this study and reflects on how these objectives have been met at different stages of the research. It then discusses the multi-dimensional model for determining susceptibility to Unintentional insider threats by examining the resulting structural model and results of the hypothesis testing.

5.2 Discussion of the Research Objectives

The main objective of this research was to develop and validate a unified and multi-dimensional theoretical model for determining susceptibility to the Unintentional Insider Threat in information systems security. In order to achieve this research objective, three specific objectives were outlined.

The first specific objective focused on establishing a theoretical foundation for the factors that contribute to the unintentional insider threat to information systems security. Prior research has highlighted a deficiency in the theoretical grounding and understanding of the Unintentional Insider Threat phenomenon (Luo et al., 2013; Tetri & Vuorinen, 2013; Wang et al., 2012; Workman, 2007). Wang et al. (2012) states that "there is great need for research that investigates the theoretical underpinning" of Unintentional insider threats. Luo et al. (2013) states that there is "limited theory-grounded research" around this phenomenon. The Carnegie Mellon University Insider Threat Team CERT (2013) present a prominent foundational report on the Unintentional Insider Threat phenomenon and call for future research to focus on causal factors. Previous studies have mostly been of an empirical nature without the required theoretical grounding. Tetri & Vuorinen (2013) point out that only 5 of the 40 articles they reviewed were studies that had analyzed some kind of empirical data. Worse still, only two of these five were explicitly grounded on theory. The literature review undertaken in Chapter 2 of this research showed similar trends. Of the 75 studies reviewed, only 21 (28%) were explicitly grounded on theory.

The first specific objective was achieved in Chapter 2 and Chapter 3. Chapter 2 identified various causal factors from extant literature and previous studied. Chapter 3 presented the unified conceptual model grounded in theory. A detailed justification for the choice of theoretical foundation and construct selection for inclusion in the model was given. The theories selected were: the Elaboration Likelihood Model by Petty & Cacioppo (1986) and the Protection Motivation Theory by Rogers (1975, 1983) with reference to its application in the Technology Threat Avoidance Theory by Liang & Xue (2009, 2010). Constructs relating to knowledge and organizational factors were added to the model based on empirical findings. They could not be ignored because empirical studies had demonstrated their causal impact on unintentional insider threat susceptibility.

The second specific objective was to develop a unified multi-dimensional theoretical model that explains susceptibility to the unintentional insider threat to information systems security. Previous studies have taken a piecemeal approach by focusing on specific aspects of the Unintentional Insider Threat phenomenon. Vishwanath et al. (2011) point out that this focus on specific causative factors does not bring out combined effects of an integrated model.

A summary of various causal factors identified through theory and review of extant literature was presented in Table 2. This table showed in a visual way how existing studies had not examined the phenomenon holistically but instead had focused on specific aspects of the phenomenon. For example, Luo et al. (2013) and Wang et al. (2012) focus on cognitive processing but do not examine organizational factors nor protection motivation through threat and coping appraisal. Arachchilage & Love (2013) and Liang & Xue (2009, 2010) on the other hand focus on threat detection and threat avoidance but do not also examine organizational factors nor do they explore cognitive factors that influence susceptibility.

Tetri & Vuorinen (2013) explicitly call for a multidimensional theoretical approach for the study of Unintentional insider threats because this provides a more holistic and explicit understanding of the phenomenon. They state that the phenomenon should not only be addressed by looking at the victim (human weakness) but should also examine the organization factors and the attacker’s tactics. This study has incorporated these aspects in the model to provide a multidimensional approach to understanding the phenomenon.

Therefore, in fulfilment of the second objective, this study presented a unified multi-dimensional model in Chapter 3 that integrates all relevant causal factors together and evaluates how they interplay and affect each other. This model was summarized graphically in Figure 8 showing how the causal relationships and theorized hypothesis.

The third and final specific objective of this research was to validate the unified multi-dimensional theoretical model using empirical data and appropriate statistical methods. Despite the conceptual presentation of such a unified theoretical and multi-dimensional model being a major contribution to the existing body of knowledge, it was important that the model be validated and tested using empirical data. Wang et al. (2012) call for a rigorous empirical validation of theory relating to the unintentional insider threat phenomenon. Vishwanath et al. (2011) present Structural Equation Modeling as the appropriate mechanism of testing such an integrated model because it allows for all constructs to be examined at the same time.

In fulfilment of the third specific objective, an empirical study was designed as described in Chapter 4. It was important to design the empirical study following the right ontological and epistemological philosophies that enable the model's validation. Justification was given for the selection of a realist, positivist, objective and deductive research approach. This kind of approach allowed for a quantitative research that is appropriate for model validation and testing.

In addition, a naturalistic field study using staged attacks that mimic real-world unintentional insider threats was used to collect data based on guidance from previous studies (Bakhshi et al., 2009; Finn & Jakobsson, 2007; Huber et al., 2009; Vishwanath et al., 2011). It was determined that this would be the best way to collect data that had high ecological validity allowing for the model testing results to be generalized to wider contexts with similar real-world settings. The use of realistic naturalistic field studies has been a challenge in the study of the unintentional insider threat phenomenon. It is difficult to get organizations willing to allow such a study on their insiders to be conducted and even for such results to be publicized.

Since many constructs could not be measured through observing insider behaviour only, a questionnaire survey was also used to collect the data for the study. Questionnaires have been used in a majority of the studies of the unintentional insider threat phenomenon and have been found an effective way of collecting data.

Finally, the Structural Equation Modeling process was used to validate and test the unified and multi-dimensional theoretical model. The analysis procedures results are detailed in Chapter 5 and are discussed next.

5.3 Discussion of the Structural Equation Modelling Process

The unified and multi-dimensional theoretical model for determining susceptibility to the Unintentional Insider Threat in information systems security was validated using the Structural Equation Modeling process. The initial conceptual model outlined in Chapter 3 and illustrated in Figure 29 consisted of 22 independent variables, 1 dependent variable and 12 demographic variables.

The dataset used to validate the model was collected using two methods. The first was through direct observations of insider interaction with a naturalistic field study using staged phishing attacks that mimicked real-world threats. The second was through questionnaire survey responses filled in by respondents who interacted with the naturalistic field study scenario. After the necessary processes of data entry, coding and screening; a total of 192 valid and usable cases were captured in the dataset and these were used in the subsequent data analysis.

An Exploratory Factor Analysis was conducted to identify the latent factors from the large number of indicators used by the measurement tools. This step was very useful in confirming that the measurement items were actually measuring the intended factors. It was also a powerful confirmation of the validity of the measurement scales that were used in data collection. This is because the resulting indicator-factor structure was unveiled by the collected dataset and not pre-determined by the model.

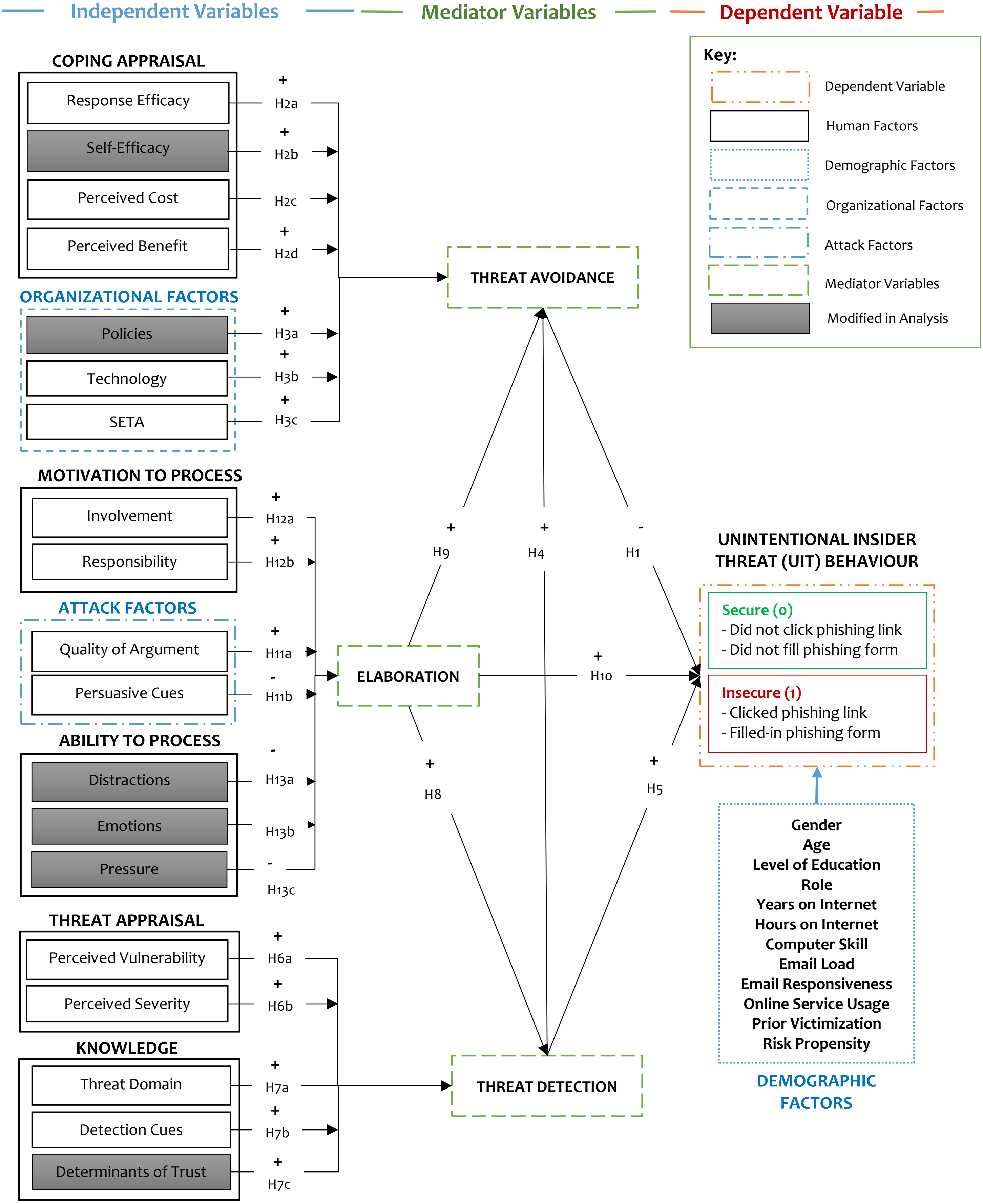


Figure 29: Changes to the Initial Multi-Dimensional Theoretical Causal Model

A few changes were made to this initial conceptual model based on the results of the Exploratory Factor Analysis. The EFA extracted 20 model constructs from 83 measurement items. This meant that three factors (Self-efficacy, Policies and Emotions) were dropped from the original theorized model. In addition, the Trust Determinants factor was split into two based on two dimensions emerging from the factor analysis. Furthermore, two factors (Distractions and Pressure) were combined into one factor indicating that the measurement items were effectively measuring one underlying construct. This latent construct was named 'Ability to Process' in line with the theoretical model.

The final EFA factor solution explained 68.78% of the total variance, it's Kaiser-Meyer-Olkin (KMO) Measure of Sampling Adequacy (MSA) was 0.761 (above the required threshold of 0.7) and the Bartlett's Test of Sphericity was statistically significant as prescribed by Hair et al. (2009). In addition, the goodness-of-fit test using Chi-square/df ratio (χ^2/df) was 1.197 which is within the desired threshold of between 1 and 3. All these parameters indicated a satisfactory indicator-factor structure that could then be used in the Confirmatory Factor Analysis.

The Confirmatory Factor Analysis is considered a vital stage of the Structural Equation Modeling process. It used to validate the measurement model before the theorized relationships can be tested. Unlike the Exploratory Factor Analysis that extracted the factor structure from the dataset without a pre-defined blueprint, the Confirmatory Factor Analysis started from the theorized model structure and examined if the hypothesized factors could be confirmed by the dataset based on the validity of the measurement model. The Confirmatory Factor Analysis saw seven indicators being dropped (PC2, KQ6, KQ3, KQ4, DIST2, DIST1, PRES3) in order to achieve satisfactory factor unidimensionality, reliability, convergent validity, discriminant validity and model fit.

The structural model analysis was performed done once the measurement model was assessed and found satisfactory. The results of the structural model analysis and hypotheses testing were illustrated in Figure 28 and are subsequently discussed in the next section

5.4 Discussion of the results of Hypotheses Testing

This section discusses each of the results obtained from the hypotheses testing through Structural Equation Modeling. There were 13 sets of hypotheses in the initial conceptual model presented in Chapter 3 and each of these is discussed hereafter.

5.4.1 Hypothesized effect of Threat Avoidance on the Unintentional Insider Threat Behaviour Outcome

The threat avoidance factor is defined as the motivation to evade an unintentional insider threat. It can be thought of as the behavioural intention before the actual behavioural outcome is manifested. A higher motivation to avoid the threat should translate to a lower unintentional insider threat behavioural outcome. The results of the hypothesis testing for this study are outlined in Table 54.

Table 54: Hypothesized effect of Threat Avoidance on Unintentional Insider Threat Behaviour Outcome

H1: Threat Avoidance has a negative and significant effect on the Unintentional Insider Threat Behaviour Outcome

Hypothesis	Path	β values	S.E.	p-value	Conclusion
H1	TAV -> UITB	0.084	0.033	0.231	Not Supported

The results are contrary to the results obtained by Arachchilage & Love (2013) and Liang & Xue (2010) where the Threat Avoidance (termed as Avoidance Motivation) had a positive and significant effect on Avoidance Behaviour. The results obtained show a positive but non-significant effect on the Unintentional Insider Threat Behaviour outcome.

A possible explanation for this could be that one of the key antecedents (Self-efficacy) was dropped during Exploratory Factor Analysis. In addition, some of the key antecedents relating to organizational factors (Technology and Security Education Training and Awareness) did not have large or significant contributions to the construct.

5.4.2 Hypothesized effect of Coping Appraisal on Threat Avoidance

The Coping Appraisal construct is from the Protection Motivation Theory by Rogers (1975, 1983) and it measures an individual's perception of: (1) response efficacy – which is the effectiveness of the recommended protective response; (2) self-efficacy – which is their ability to execute the recommended protective response; (3) perceived cost and (4) perceived benefit of recommended protective responses. Each of these four factors was captured as sub-hypothesis and the results of the hypothesis testing are as outlined in Table 55.

Table 55: Hypothesized effect of Coping Appraisal on Threat Avoidance

H2a: Response Efficacy has a positive and significant effect on Threat Avoidance

H2b: Self-Efficacy has a positive and significant effect on Threat Avoidance

H2c: Perceived Response Cost has a negative and significant effect on Threat Avoidance

H2d: Perceived Response Benefit has a positive and significant effect on Threat Avoidance

Hypothesis	**Path**	**β values**	**S.E.**	**p-value**	**Conclusion**
H2a	RE -> TAV	0.306	0.114	***	Supported
H2b	SE -> TAV	-	-	-	Dropped at EFA
H2c	RC -> TAV	-0.155	0.087	0.046	Supported
H2d	PB -> TAV	0.089	0.134	0.225	Not Supported

Liang & Xue (2010) tested the effect of Response Efficacy (termed as Safeguard Effectiveness), Self-efficacy and Response Cost (termed as Safeguard Cost) on Threat Avoidance (termed as Avoidance Motivation). Results of their study showed that Response Efficacy had the highest effect on Threat Avoidance (0.33) and this effect was positive and significant at $p \leq .01$ level. Self-Efficacy had the second highest effect on Threat Avoidance (0.19) and this effect was positive and significant at $p \leq .05$ level. Finally, Response Cost had a negative (-0.14) and significant effect at $p \leq .05$ level on Threat Avoidance.

Lee & Larsen (2009) also examined the effect of Response Efficacy, Self-Efficacy and Response Cost on Small-and-Medium sized Business executives' intention to adopt anti-malware software for their companies. They found that Response-Cost had the highest effect (-0.257 significant at $p \leq .001$ level), followed by Response Efficacy (0.215 significant at $p \leq .001$ level) and finally Self-Efficacy (0.114 significant at $p \leq .05$ level).

Similar to these prior studies, results obtained showed a highly significant positive effect from Response Efficacy (0.306 significant at $p \leq .001$ level) and significant negative effects from Response cost (-0.155 significant at $p \leq .05$ level). Response Efficacy had the highest effect on Threat Avoidance (0.306) followed by Response Cost. However, unlike previous studies, Perceived Benefit did not have a significant effect; though it was in the hypothesized direction. A possible explanation for the non-significant effect is that the Response Cost considerations outweigh the

Response Benefit provisions when individuals are considering threat avoidance strategies. The Self-Efficacy construct was dropped at Exploratory Factor Analysis due to measurement item inadequacies.

5.4.3 Hypothesized effect of Organizational Factors on Threat Avoidance

The incorporation of factors relating to organizational defenses was informed from best-practice frameworks such as ISO27000 series, COBIT and NIST Special Publications in Information Security which advocate for the use of policy, technology and security education training and awareness as controls to mitigate information security threats. Each of these organizational factors was specified as a sub-hypothesis and results of the hypothesis testing are as outlined in Table 56.

Table 56: Hypothesized effect of Organizational Factors on Threat Avoidance

H3a: Policies have a positive and significant effect on Threat Avoidance

H3b: Technology Controls have a positive and significant effect on Threat Avoidance

H3c: Security Education Training and Awareness have a positive and significant effect on Threat Avoidance

Hypothesis	Path	β values	S.E.	p-value	Conclusion
H3a	POL -> TAV	-	-	-	Dropped at EFA
H3b	TC -> TAV	0.031	0.058	0.665	Not Supported
H3c	SETA -> TAV	0.091	0.054	0.209	Not Supported

Policy was dropped at the Exploratory Factor Analysis level and the other two factors were not found to have significant effects on Threat Avoidance. They however showed that the hypothesized direction was correct. This means that incorporating technology controls and security education training and awareness programs would increase chances of threat avoidance.

It is important to note that previous empirical studies that were reviewed in Chapter 2 in the area of unintentional insider threats did not incorporate organizational defenses in their hypothesis testing. Previous studies did however show that giving individuals security education training and awareness did reduce their susceptibility to unintentional insider threats (Kumaraguru et al., 2009, 2008; Sheng et al., 2010). In addition the use of technology controls, such as browser and e-mail based extensions and warnings, also reduces susceptibility to unintentional insider threats (Downs et al., 2007; Egelman et al., 2008). None of the studies that were reviewed examined the effect of security policies on susceptibility to unintentional insider threats however Tetri & Vuorinen (2013) did recommend for future studies to examine their effect.

5.4.4 Hypothesized effect of Threat Detection on Threat Avoidance

Threat detection was conceptualized as the extent to which an individual is able to correctly perceive a danger. Previous studies by Arachchilage & Love (2013) and Liang & Xue (2010) examined the effect of threat detection (using the term perceived threat) on threat avoidance (termed as avoidance motivation). Arachchilage & Love (2013) found that threat detection has a positive (0.39) and significant effect (at $p \leq .01$ level) on threat avoidance. Similarly, Liang & Xue (2010) found a positive (0.26) and significant effect (at $p \leq .01$ level) of threat detection on threat avoidance. This means that if an individual can identify a threat, they are expected to have an intention and motivation to avoid the threat. The results of the hypothesis testing for this study are outlined in Table 57.

Table 57: Hypothesized effect of Threat Detection on Threat Avoidance

H4: Threat Detection has a positive and significant effect on Threat Avoidance

Hypothesis	Path	β values	S.E.	p-value	Conclusion
H4	TD -> TAV	0.115	0.054	0.129	Not Supported

The results of the hypothesis testing were not significant and the hypothesis was not supported by the dataset. This could be because the effect of Threat Detection was not strong enough to be significant. This could also be because the key antecedents relating to Threat Appraisal (Perceived Vulnerability and Perceived Severity) were also neither strong nor significant.

5.4.5 Hypothesized effect of Threat Detection on the Unintentional Insider Threat Behaviour Outcome

Instead of examining the effect of Threat Detection though the mediated effect of Threat Avoidance, this study examined the direct effect of Threat Detection on the Unintentional Insider Threat behavioural outcome. This is unlike the studies by Arachchilage & Love (2013) and Liang & Xue (2010) that only examined the mediated effect of Threat Detection through Threat Avoidance. The results of the hypothesis testing for this study are outlined in Table 58.

Table 58: Hypothesized effect of Threat Detection on the Unintentional Insider Threat Behaviour Outcome

H5: Threat Detection has a negative and significant effect on the Unintentional Insider Threat Behaviour Outcome					
Hypothesis	**Path**	**β values**	**S.E.**	**p-value**	**Conclusion**
H5	TD -> UITB	-0.392	0.024	***	Supported

The results show that Threat Detection has a negative (-0.392) and significant effect (at p ≤ .001 level) on the Unintentional Insider Threat behavioural outcome. This means that an individual with high threat detection is less likely to become susceptible to Unintentional Insider Threats.

5.4.6 Hypothesized effect of Threat Appraisal on Threat Detection

The Threat Appraisal construct is borrowed from the Protection Motivation Theory by Rogers (1975, 1983). It represents the extent to which an individual correctly judges the risk of harm they face due to a threat. Two components inform the Threat Appraisal construct; the Perceived Severity (level of harm that the threat could cause) and the Perceived Vulnerability (level an individual is exposed to the possibility of succumbing to the threat).

Previous studies by Arachchilage & Love, 2013; Liang & Xue, 2010; Workman, 2007 and Workman et al., 2008 have examined the effect that Perceived Severity and Perceived Vulnerability have on the susceptibility to unintentional Insider Threats. Arachchilage & Love (2013) and Liang & Xue (2010) have specifically shown that the effects of Perceived Severity and Perceived Vulnerability are mediated by Threat Detection (termed as Perceived Threat in their work).

In their study Arachchilage & Love (2013) found that Perceived Severity had a higher effect (0.50 significant at p ≤ .01 level) on threat perception than Perceived Vulnerability (0.36 significant at p ≤ .01 level). On the contrary, Liang & Xue (2010) found that Perceived Vulnerability had a higher effect (0.41 significant at p ≤ .01 level) as compared to Perceived Severity (0.27 significant at p ≤ .01 level).

The results of the hypothesis testing for this study are outlined in Table 59.

Table 59: Hypothesized effect of Threat Appraisal on Threat Detection

<table>
<tr><td colspan="6"><b>H6a</b>: Perceived Vulnerability has a positive and significant effect on Threat Detection</td></tr>
<tr><td colspan="6"><b>H6b</b>: Perceived Severity has a positive and significant effect on Threat Detection</td></tr>
<tr><th>Hypothesis</th><th>Path</th><th>β values</th><th>S.E.</th><th>p-value</th><th>Conclusion</th></tr>
<tr><td><b>H6a</b></td><td>PVUL -> TD</td><td>0.002</td><td>0.077</td><td>0.973</td><td>Not Supported</td></tr>
<tr><td><b>H6b</b></td><td>PS -> TD</td><td>0.025</td><td>0.14</td><td>0.685</td><td>Not Supported</td></tr>
</table>

The results show that neither Perceived Vulnerability nor Perceived Severity had a strong effect on Threat Detection and these effects were not significant. A possible explanation for this is that Threat Detection fully mediated the effects of these two factors.

5.4.7 Hypothesized effect of Knowledge on Threat Detection

Knowledge is defined as the information and skills an individual acquires that affects their understanding of a matter. Previous studies have shown that an individual's knowledge allows the individual to correctly perceive threats when they present themselves (Dhamija et al., 2006; Downs et al., 2006; Fogg et al., 2001; Friedman et al., 2002; Garera et al., 2007; Grazioli, 2004; Jakobsson & Ratkiewicz, 2006; Jakobsson et al., 2007; Karakasiliotis et al., 2006; Sheng et al., 2010; Tsow & Jakobsson, 2007; Vishwanath et al., 2011; Wang et al., 2012).

This research empirically analyzed the effect of three categories of Knowledge in order to determine which specific types of knowledge have an impact on threat detection. Knowledge on the information security threat domain, on detection cues and on trust determinants were each examined as separate sub-hypothesis. The results of the hypothesis testing for this study are outlined in Table 60.

Table 60: Hypothesized effect of Knowledge on Threat Detection

<table>
<tr><td colspan="6"><b>H7a:</b> Knowledge on Threat Domain has a positive and significant effect on Threat Detection</td></tr>
<tr><td colspan="6"><b>H7b:</b> Knowledge on Detection Cues has a positive and significant effect on Threat Detection</td></tr>
<tr><td colspan="6"><b>H7c:</b> Knowledge on Low Determinants of Trust has a negative and significant effect on Threat Detection</td></tr>
<tr><td colspan="6"><b>H7d:</b> Knowledge on High Determinants of Trust has a positive and significant effect on Threat Detection</td></tr>
<tr><th>Hypothesis</th><th>Path</th><th>β values</th><th>S.E.</th><th>p-value</th><th>Conclusion</th></tr>
<tr><td><b>H7a</b></td><td>KQC -> TD</td><td>0.033</td><td>0.235</td><td>0.63</td><td>Not Supported</td></tr>
<tr><td><b>H7b</b></td><td>DC -> TD</td><td>0.468</td><td>0.066</td><td>***</td><td>Supported</td></tr>
<tr><td><b>H7c</b></td><td>DT_LOW -> TD</td><td>-0.286</td><td>0.122</td><td>***</td><td>Supported</td></tr>
<tr><td><b>H7d</b></td><td>DT_HIGH -> TD</td><td>0.276</td><td>0.069</td><td>***</td><td>Supported</td></tr>
</table>

The results for the first type of knowledge showed that knowledge on the information security threat domain did not have a significant effect on threat detection. This could be because Threat Detection fully mediates its effect on the Unintentional Insider Threat Behavioural outcome. The positive direction of the effect (0.033) shows that an increase in knowledge on the threat domain does increase the chances of threat detection. The results on the second type of knowledge showed that knowledge on detection cues has a positive effect (0.468) on threat detection and this effect is highly significant at the $p \leq .001$ level. This means that as individuals increase in their understanding of cues that can help them identify unintentional insider threats, they also increase in their ability to detect the threats.

During the Exploratory Factor Analysis knowledge on the third category of information (Determinants of Trust) has two distinct dimensions. The first is low determinants of trust. Fogg et al. (2001); Jakobsson et al. (2007) and Tsow & Jakobsson, (2007) in their studies showed that the use of grammar, spelling, look and feel and personalized messages increased user's perceptions of credibility. However, these characteristics are easily manipulated by attackers to make their messages deceptive. In fact Dhamija et al. (2006) showed that users fall for deceptive phishing messages because they pay attention to such determinants of trust. The results of this study confirm these previous findings and show that the individuals who relied on consistency in logo, colours, look and feel, grammar and spelling, personalized greetings with their names, reasonableness of content and context and even email address of the sender were less able to detect the threat. The direction of the effect (-0.286) shows a negative relationship between the use of low determinants of trust and threat detection and this effect is highly significant at the $p \leq .001$ level.

However, Dhamija et al. (2006) also showed that users who had proper knowledge and understanding about security indicators such as address and status bar indicators, SSL certificates and chrome padlock icons were less susceptible to deception. The results of this study confirm these previous findings and show that individuals who rely on web address and hyperlink evaluation, website encryption or padlock icon, website certificate, website registration information and security tool information are less likely to fall for unintentional insider threats. The effect is positive (0.276) and is highly significant at the $p \leq .001$ level.

5.4.8 Hypothesized effect of Elaboration on Threat Detection

The Elaboration construct is derived from the Elaboration Likelihood Model (ELM) by Petty & Cacioppo (1986) and it represents the extent to which an individual cognitively evaluates a persuasive message by particularly paying attention to the issue-relevant arguments as opposed to distractive persuasive cues. Wang et al. (2012) showed that an increase in cognitive effort evaluating a deceptive message reduces the likelihood to respond to the message. The hypothesized negative effect was supported in their findings (-0.026) but this effect was not found to be significant. Their research examined the direct relationship between Elaboration and the outcome behaviour. This study however also examines the effect of Elaboration on Threat Detection. The results of the hypothesis testing are outlined in Table 61.

Table 61: Hypothesized effect of Elaboration on Threat Detection

<table>
<tr><td colspan="6"><b>H8</b>: Elaboration has a positive and significant effect on Threat Detection</td></tr>
<tr><th>Hypothesis</th><th>Path</th><th>β values</th><th>S.E.</th><th>p-value</th><th>Conclusion</th></tr>
<tr><td><b>H8</b></td><td>ELAB -> TD</td><td>0.229</td><td>0.058</td><td>***</td><td>Supported</td></tr>
</table>

The results of this study show that indeed an increase in Elaboration also increases Threat Detection and this positive effect (0.229) is highly significant at the $p \leq .001$ level.

5.4.9 Hypothesized effect of Elaboration on Threat Avoidance

This study also examines the effect Elaboration has on Threat Avoidance. The results of the hypothesis testing are outlined in Table 62.

Table 62: Hypothesized effect of Elaboration on Threat Avoidance

<table>
<tr><td colspan="6"><b>H9</b>: Elaboration has a positive and significant effect on Threat Avoidance</td></tr>
<tr><th>Hypothesis</th><th>Path</th><th>β values</th><th>S.E.</th><th>p-value</th><th>Conclusion</th></tr>
<tr><td><b>H9</b></td><td>ELAB -> TAV</td><td>0.183</td><td>0.051</td><td>0.015</td><td>Supported</td></tr>
</table>

The results show that as elaboration increases, the threat avoidance also increases. This positive effect (0.183) is also significant at the $p \leq .05$ level.

5.4.10 Hypothesized effect of Elaboration on the Unintentional Insider Threat Behaviour Outcome

The study by Wang et al. (2012) hypothesized that an increase in cognitive effort evaluating a deceptive message reduces the likelihood to respond to the message.

The results of their study showed a negative effect (-0.026) but this effect was not found to be significant. The results of the hypothesis testing for this study are in Table 63.

Table 63: Hypothesized effect of Elaboration on the Unintentional Insider Threat Behavioural Outcome

H10: Elaboration has a negative and significant effect on the Unintentional Insider Threat Behavioural Outcome					
Hypothesis	**Path**	**β values**	**S.E.**	**p-value**	**Conclusion**
H10	ELAB -> UITB	-0.018	0.022	0.8	Not Supported

The results are similar to what Wang et al. (2012) found. The effect was in the hypothesized direction (-0.018) but was not significant. This could be because the effects of Elaboration on the Unintentional Insider Threat behaviour are fully mediated by Threat Detection and Threat Avoidance.

5.4.11 Hypothesized effect of Attack Factors on Elaboration

In their study, Tetri & Vuorinen (2013) recommend that future research on unintentional insider threats consider the role of the intruder (attacker) on influencing susceptibility. The Elaboration Likelihood Model (ELM) by Petty & Cacioppo (1986) present two constructs that can be attributed to the quality of the attack as designed by the attacker. These are the use of persuasive cues and also the quality of argument.

The study by Luo et al. (2013) hypothesized that individuals are likely to be more victimized by messages with high argument quality. Results of their exploratory study showed that well-crafted messages led to more victimization.

In contrast, Wang et al. (2012) hypothesized that attention to persuasive cues (what they termed visceral triggers) led to reduced elaboration and their results supported this with a negative effect (-0.179) that was significant at $p \leq .05$ level.

The results of the hypothesis testing for this study are outlined in Table 64.

Table 64: Hypothesized effect of Attack Factors on Elaboration

H11a: Quality of Argument has a positive and significant effect on Elaboration					
H11b: Persuasive Cues have a negative and significant effect on Elaboration					
Hypothesis	**Path**	**β values**	**S.E.**	**p-value**	**Conclusion**
H11a	QA -> ELAB	0.622	0.07	***	Supported
H11b	PC -> ELAB	-0.021	0.072	0.733	Not Supported

The results of this study showed that the higher the quality of argument, the higher the level of elaboration. This effect was in the hypothesized direction (0.622) and was highly significant at the p ≤ .001 level. This was in line with the study by Luo et al. (2013).

The results also showed that attention to persuasive cues reduced the level of elaboration. This effect (-0.021) was in the negative direction as hypothesized but its effect was not significant. This was different from the results by Wang et al. (2012) that showed significant negative effects. This could have been because the insiders paid more attention to the argument as opposed to the persuasive cues used in the study.

5.4.12 Hypothesized effect of Motivation to Process on Elaboration

The Elaboration Likelihood Model (ELM) by Petty & Cacioppo (1986) present the 'Motivation to Process' construct. They explain that an individual's motivation to process a deceptive message will be influenced by their level of involvement and also responsibility.

Previous studies by Vishwanath et al. (2011) and Wang et al. (2012) examined the effect of Involvement on Elaboration but they did not study the effect of Responsibility. Vishwanath et al. (2011) found that Involvement had a positive effect on Elaboration (0.20) and this effect was significant at the p ≤ .05 level. Similarly, Wang et al. (2012) found a positive effect of involvement on Elaboration (0.178) with a higher level of significance at the p ≤ .01 level.

The results of the hypothesis testing for this study are outlined in Table 65.

Table 65: Hypothesized effect of Motivation to Process on Elaboration

H12a: Involvement has a positive and significant effect on Elaboration					
H12b: Responsibility has a positive and significant effect on Elaboration					
Hypothesis	**Path**	**β values**	**S.E.**	**p-value**	**Conclusion**
H12a	INV -> ELAB	-0.092	0.063	0.14	Not Supported
H12b	RES -> ELAB	0.133	0.102	0.043	Supported

The results of this study show that involvement had a negative effect on elaboration, contrary to what was hypothesized or shown in previous studies. This effect however was not significant. It could be because the insiders with a sense of involvement were not as motivated to think too much about the persuasive message but to just accept it.

This study studies the effect of responsibility unlike the previously mentioned studies that only examined involvement. The results of the hypothesis testing showed that responsibility had a positive effect on elaboration; meaning the more an individual was responsible for the issues relating to the persuasive message the more they would cognitively evaluate the message. The results also showed that the effect (0.133) was also significant at the p ≤ .05 level.

5.4.13 Hypothesized effect of Ability to Process on Elaboration

The Elaboration Likelihood Model (ELM) by Petty & Cacioppo (1986) also presents the ‘Ability to Process’ construct. This study not only examined the effect that Distractions have on Elaboration but also examined the effect that Emotions and Pressure have on Elaboration. Distractions are identified from the Elaboration Likelihood Model as having a negative effect on the ability to process. The effect of Emotions and Pressure on the Ability to Process are borrowed from Cialdini's (2001) six principles of influence and persuasion. Luo et al. (2013) in their study hypothesize that Distractions and Time Pressure suppress Elaboration.

The three different dimensions (Distractions, Emotions and Pressure were dropped during the Exploratory Factor Analysis and a unifying underlying latent factor named “Ability to Process” was studied instead. The results of the hypothesis testing for this study are outlined in Table 66.

Table 66: Hypothesized effect of Ability to Process on Elaboration

H13: Ability to Process has a positive and significant effect on Elaboration					
H13a: Distractions have a negative and significant effect on Elaboration					
H13b: Emotions have a negative and significant effect on Elaboration					
H13c: Pressure has a negative and significant effect on Elaboration					
Hypothesis	**Path**	**β values**	**S.E.**	**p-value**	**Conclusion**
H13	AP -> ELAB	0.019	0.114	0.785	Not Supported
H13a	DIST -> ELAB	-	-	-	Removed
H13b	EM -> ELAB	-	-	-	Removed
H13c	PRES -> ELAB	-	-	-	Removed

The results of this study showed that an individual with a high Ability to Process (that is not impaired by distractions, emotions or pressure) has a high level of Elaboration. The results confirmed this positive effect of the ability to process on elaboration (0.019) however this effect was neither strong nor significant. It could be because the "Ability to Process" construct was not well captured since a number of measurement items were dropped during the factor analysis process.

5.5 Discussion of the effect of Demographic Factors on the Unintentional Insider Threat Behaviour Outcome

A total of twelve demographic variables were examined in this study to see their effect on the Unintentional Insider Threat behavioural outcome. These variables were set aside as demographic variables because they were not considered theoretical constructs in the model. It was therefore important to observe the effects of the model constructs while controlling for them. The demographic variables examined were: Gender, Age, Level of Education, Role, Years on the Internet, Hours on the Internet, Computer Skill, Email Load, Email Responsiveness, Online Service Usage, Prior Victimization and Risk Propensity. The effects of the demographic variables for this study are shown in Table 67.

Table 67: Effect of Demographic Variables on Unintentional Insider Threat Behaviour Outcome

Demographic Variable	Path	β values	S.E.	p-value	Conclusion
Gender	GENDER -> UITB	-0.029	0.055	0.649	Not significant
Age	AGE -> UITB	0.119	0.022	0.062	Marginally significant
Level of Education	EDUCATION -> UITB	-0.061	0.023	0.343	Not significant
Role	ROLE -> UITB	0.189	0.03	0.003	Significant
Years on Internet	YEAR_INTERNET -> UITB	0.145	0.023	0.024	Significant
Hours on Internet	HOURS_INTERNET-> UITB	0.072	1.118	0.264	Not significant
Computer Skill	COMP_SKILL -> UITB	-0.095	0.032	0.137	Not significant
Email Load	EL -> UITB	-0.157	0.021	0.014	Significant
Email Responsiveness	ER1 -> UITB	0.128	0.022	0.046	Significant
Online Service Usage	OS1 -> UITB	-0.039	0.043	0.539	Not significant
Prior Victimization	PV1 -> UITB	-0.019	0.074	0.761	Not significant
Risk Propensity	RP -> UITB	0.02	0.03	0.774	Not significant

The results show that Age, Role, Years on the Internet, Email Load and Email Responsiveness have a significant effect on the Unintentional Insider Threat behaviour outcome but the other demographic variables do not.

Gender was not found to have a statistically significant effect on the Unintentional Insider Threat behaviour similar to the findings of Arachchilage & Love (2013) and Wang et al. (2012). This is in contrast to previous studies by Jagatic et al. (2007) and Sheng et al. (2010) that showed that women were more susceptible to unintentional Insider Threats than men. When they did further analysis, they found that this was because women had less technical knowledge and training than men. It could be that this knowledge-gap across the genders has been addressed over time and therefore recent studies do not reflect this gender difference.

Age was found to have a marginally significant effect on the Unintentional Insider Threat behavioural outcome. The study by Wang et al. (2012) found that Age had a negative effect (-0.145) and this effect was significant at $p \leq .001$ level. This meant that older individuals were less susceptible to the Unintentional Insider Threat behaviour. However, this study found this effect to be positive (0.119) and marginally significant at $p \leq 0.1$ level. This study indicates that younger people are less susceptible to the unintentional insider threats. The results of exploratory cluster analysis shows that the younger 18-25 age group is comprised of undergraduate students who seem to be familiar with the kind of Unintentional Insider Threat scenario staged and therefore less susceptible as compared to the older counterparts.

Level of Education was not found to have a significant effect on the Unintentional Insider Threat behavioural outcome. This is unlike the study by Sheng et al. (2010) that found the level of education to be more influential than age or gender. The reason for this could be that the knowledge construct included in the theoretical model addresses the effect regardless of the level of education of an individual.

Role was found to have a positive effect on the Unintentional Insider Threat behavioural outcome and this effect was significant at $p \leq 0.05$ level. Three roles were examined in the study: student, faculty and staff. This meant that students were less susceptible to the unintentional insider threat as compared to other roles. This is unlike the study by Kumaraguru et al. (2009) that found students to be most susceptible. The students in this study could be more tech-savvy and more familiar with the threat scenario making them less susceptible. Another explanation could be that they were less engaged because they rarely use their official email accounts and the staged scenario may not have had the same effect as for the faculty and staff who must operate their official email accounts because of their work.

Years on the internet was found to have a significant effect on the Unintentional Insider Threat behavioural outcome. This effect was positive unlike the study by Sheng et al. (2007) that showed that people who have more years of experience on the internet are able to detect threats better and are therefore less susceptible. The results of this study could point to a need to qualify the type of experience gained from long term usage of the internet. It could be that the years on the internet do not necessarily build on the kind of knowledge needed to avoid unintentional insider threats especially if such knowledge is not passed on deliberately over the internet.

Similarly, hours on the internet had a positive effect on the unintentional insider threat behavioural outcome but this effect was not found to be significant. Previous studies by Fogg et al. (2001) and Arachchilage & Love (2013) studied this variable using the term "internet experience" and they also did not establish that it had a significant effect on susceptibility to unintentional insider threats. The positive effect captured in the results could mean that individuals who spend many hours on the internet are also more exposed to unintentional insider threats and therefore more likely to succumb to the threats. It could also mean that increased hours on the internet does not directly translate to more knowledge on how to detect and avoid such threats.

Computer skill was not found to have a significant effect on the unintentional insider threat behavioural outcome. The effect was however negative (-0.095) as would be expected based on previous studies by Jagatic et al. (2007), Kumaraguru et al. (2009) and Sheng et al. (2010). These previous studies have shown that computer-savvy individuals are less susceptible to unintentional insider threats.

Email load was found to have a significant effect on the unintentional insider threat behavioural outcome and this effect was negative. A previous study by Vishwanath et al. (2011) showed that the more emails a person received a day, the more likely they were to succumb to phishing attacks. This study however showed that the people who received many emails were less likely to fall victim to unintentional insider threats. A possible explanation for this is that the individuals could have many emails and therefore may not process all the emails and therefore they may not give much attention to suspicious emails. It could also be that they have seen many such threats and they have learnt to ignore them.

On the other hand, Email Responsiveness was found to have a positive effect (0.128) that was significant (at $p \leq 0.05$ level) on the unintentional insider threat behavioural outcome. This was more in line with the study by Vishwanath et al. (2011); meaning it is not the fact that people receive more emails that makes them susceptible it is whether they actually take the time to process and respond to these emails that matters. It could be that people who are keen to respond to their emails fall for phishing attacks because they respond mechanically and not because of consciously reasoned action. This is what Vishwanath et al. (2011) terms as the effect of habituation.

Online Service Usage was not found to have a significant effect on the unintentional insider threat behavioural outcome. The effect was however negative (-0.039) as expected based on a previous study by Downs et al. (2007) that found that participants who used many online services were less likely to click phishing links. This could be because such services intentionally make effort and take time to educate their customers against such threats.

Prior victimization was not found to have a significant effect on the unintentional insider threat behavioural outcome. However, this effect was negative (-0.019) as would be expected based on a previous study by Workman (2008b) which established that previous victims of social engineering attacks were less susceptible to later threats. Similar to this study, Workman (2008b) did not find this effect to be significant. It could be because its effects are addressed by the threat appraisal construct built into the theoretical model.

Risk Propensity was not found to have a significant effect on the unintentional insider threat behavioural outcome. However, as expected, the effect was found to be positive. This expectation is based on a previous study by Sheng et al. (2010) that showed that the more risk-averse individuals were less likely to succumb to unintentional insider threats. It may be better to measure risk propensity using a better measurement scale or tool that objectively captures an individual's risk profile such as the Balloon Analogue Risk Task (BART) method (Hunt, Hopko, Bare, Lejuez, & Robinson, 2005; Lejuez et al., 2002).

5.6 Discussion of the overall model prediction

As shown in Figure 30, the multi-dimensional model for determining susceptibility to unintentional insider threats presented in this study is able to account for 28.7% of the variance in the dependent variable ($R^2 = 0.287$). This is much better compared to the recent study by Arachchilage & Love (2013) that accounted for 15% of the variance of the behavioural outcome and also better than the model by Wang et al. (2012) that explained 16% of the variance in response likelihood. Similarly, the mediated model by Vishwanath et al. (2011) accounted for 20% of the variance in likelihood to respond. This study's model is also able to account for 43.1% of the variance in the Threat Detection endogenous explanatory variable ($R^2 = 0.431$). This is much better than the model by Arachchilage & Love (2013) that explained 36% of the perceived threat. In addition, this study's model is able to account for 19.1% of the Threat Avoidance endogenous explanatory variable ($R^2 = 0.191$). This is slightly less than that of the model by Arachchilage & Love (2013) that explained 21% of the avoidance motivation construct. Another endogenous explanatory variable that was studied was Elaboration. This study's model was able to explain 41.4% of the variance in Elaboration. This is better than the model by Wang et al. (2012) which accounted for 11% of the variance in the cognitive effort variable. The previous study by Arachchilage & Love (2013) did not examine the effect of Elaboration.

5.7 Chapter Summary

This chapter has discussed the findings from this research in the context of previous studies from extant literature. It has started by outlining the research objectives set out at the beginning of the study and discussing how each research objective has been addressed in the research. In addition, the chapter discusses the results of each of the hypothesis outlined in the multi-dimensional model. The model's overall predictive capability is also discussed showing it is a robust model compared to models presented in previous studies.

CHAPTER 6: CONCLUSIONS AND RECOMMENDATIONS

6.1 Introduction

This is the last chapter of the thesis. It aims to tie together the research findings and draw a conclusion based on the key findings that emerged from the study. This chapter also reflects on the contributions made by the research as relates to theory, existing body of knowledge, policy and practice. A critique of the strengths and limitations of the study are presented with recommendations for future research.

6.2 Summary of Findings

This study has developed and validated a unified and multi-dimensional theoretical model for determining susceptibility to the unintentional insider threat in information systems security. The model is theoretically grounded on the Elaboration Likelihood Model by Petty & Cacioppo (1986) and the Protection Motivation Theory by Rogers (1975, 1983) with reference to the Technology Threat Avoidance Theory by Liang & Xue (2009, 2010).

A review of previous studies revealed limited theory-grounded research regarding the unintentional insider threat phenomenon. In addition, most studies focused on the insider but did not examine factors relating to the organization or even the attacker to give a holistic understanding of the phenomenon. This study incorporates factors relating to the organization defenses and attack quality in order to give a more comprehensive multi-dimensional understanding of the unintentional insider threat phenomenon.

The model was validated using data from a naturalistic field experiment and using the Structural Equation Modeling technique. The entire integrated model was validated as a whole and not in piecemeal. This allowed all the 22 independent variables, 1 dependent variable and 12 demographic variables and their inter-relationships to be examined at the same time using Structural Equation Modeling.

The results of this validation provided informative insights. The overall model had a very good predictive power compared to models developed in previous studies.

The model was able to account for 28.7% of the variance in the unintentional insider threat behavioural outcome. This was better than the model by Arachchilage & Love (2013) that accounted for 15% of the variance of the behavioural outcome and also better than that by Wang et al. (2012) that explained 16% of the variance in response likelihood.

In addition, the model had three endogenous explanatory variables namely; Threat Detection, Threat Avoidance and Elaboration. It was able to account for 43.1% of the variance in Threat Detection, 19.1% of the variance in Threat Avoidance and 41.4% of the variance in Elaboration.

The results of the hypothesis testing also revealed notable findings. Five of the initial hypotheses were removed because their associated factors were dropped during the Exploratory Factor Analysis. This indicates that the measurement items developed to capture these factors were not adequate. These hypotheses related to the following factors: Self-Efficacy, Policies, Distraction, Emotions and Pressure. However, the Trust Determinants factor was split into two based on two dimensions that emerged from Exploratory Factor Analysis. Furthermore, two factors (Distractions and Pressure) were merged based on the results of the Exploratory Factor Analysis that showed they were all measuring one underlying latent factor (which was named Ability to Process).

In addition to these findings, 12 of the hypotheses were not supported by the data that was collected and analyzed in this study. These hypotheses are in relation to the following relationships: Threat Avoidance on the unintentional insider threat Behaviour; Perceived Benefit on Threat Avoidance; Technology Controls on Threat Avoidance; Security Education Training and Awareness on Threat Avoidance; Threat Detection on Threat Avoidance; Perceived Vulnerability on Threat Detection; Perceived Severity on Threat Detection; Knowledge on Threat Domain on Threat Detection; Elaboration on the Unintentional Insider Threat Behaviour; Persuasive Cues on Elaboration; Involvement on Elaboration; and Ability to Process on Elaboration.

However, 10 hypotheses were supported by the dataset collected and analyzed in this study. These hypotheses were found to have path coefficients that supported the direction of the hypothesized relationships and they also passed the tests of significance at $p \leq 0.1$, $p \leq 0.05$ and $p \leq 0.001$ levels.

The hypotheses that were supported in this study are outlined in Table 68:

Table 68: Hypotheses that were supported

Hypothesis	Path	Conclusion
H2a: Response Efficacy has a positive and significant effect on Threat Avoidance	RE -> TAV	Supported
H2c: Perceived Response Cost has a negative and significant effect on Threat Avoidance	RC -> TAV	Supported
H5: Threat Detection has a negative and significant effect on the Unintentional Insider Threat Behaviour Outcome	TD -> UITB	Supported
H7b: Knowledge on Detection Cues has a positive and significant effect on Threat Detection	DC -> TD	Supported
H7c: Knowledge on Low Determinants of Trust has a negative and significant effect on Threat Detection	DT_LOW -> TD	Supported
H7d: Knowledge on High Determinants of Trust has a positive and significant effect on Threat Detection	DT_HIGH -> TD	Supported
H8: Elaboration has a positive and significant effect on Threat Detection	ELAB -> TD	Supported
H9: Elaboration has a positive and significant effect on Threat Avoidance	ELAB -> TAV	Supported
H11a: Quality of Argument has a positive and significant effect on Elaboration	QA -> ELAB	Supported
H12b: Responsibility has a positive and significant effect on Elaboration	RES -> ELAB	Supported

These results of the hypotheses testing show that:

- Threat Detection is the most influential factor in influencing an insider’s unintentional insider threat behaviour.
- The knowledge that positively influences Threat Detection most is in relation to Threat Detection Cues and High Trust Determinants (that relate to security indicators). Insiders who rely on Low Trust Determinants (such as logo, colours, look and feel, grammar and spelling, personalized greetings, reasonableness of content and context and email address of the sender) are less able to detect threats. This is because many threats are well crafted and manipulate these low determinants of trust to provide legitimacy.
- Response Efficacy and Response Cost are the most influential factors from the Coping Appraisal construct in influencing and insider’s Threat Avoidance motivation.
- Elaboration of a threat scenario significantly influences Threat Detection. This means that the more cognitively engaged and insider is, the more likely they are to detect threats. If they do not intentionally engage themselves to evaluate deceptive messages, they are not likely to detect the threat.

- Elaboration of a threat scenario also significantly influences Threat Avoidance. This means that insiders who cognitively engage themselves to process a threat scenario are also able to determine a way to avoid the threat successfully.
- Quality of Argument is the most influential factor from the Attack Factors construct in influencing Elaboration. If an attack is made to be as believable as possible to the insider, it increases the attack quality.
- Responsibility is the most influential factor from the Motivation to Process construct in influencing Elaboration. This means that the more an insider feels responsible for the issues highlighted in the threat scenario, the more they are likely to process the threat scenario.

In addition to these theoretically defined factors, the effect of 12 demographic variables was controlled for. These variables were found to have some effect on the unintentional insider threat behaviour outcome from the evaluation of various empirical studies. Five of the demographic variables were found to have a significant effect on the unintentional insider threat behaviour outcome. These are: age, role, years on internet, email load and email responsiveness.

These results of demographic variable analysis show that:

- Older insiders are more likely to succumb to unintentional insider threats than younger ones.
- Staff and Faculty are more likely to succumb to unintentional insider threats than students.
- The more years an insider has been on the internet the more likely they are to succumb to unintentional insider threats.
- The more email load an insider has the less likely they are to succumb to unintentional insider threats.
- The more responsive an insider is to their emails the more likely they are to succumb to unintentional insider threats.

It is important to reflect on the results of the demographic variable analysis in relation to the theoretical model. For example, the effects seen on many of the demographic variables (age, role, years on internet) could be explained better by examining the knowledge an insider has on threat detection cues and trust determinants. Similarly, the effects of email load and email responsiveness could be better explained by examining the actual cognitive effort an insider expends in processing a threat scenario.

6.3 Conclusion of the Study

The primary aim of this study was to develop and validate a unified and multi-dimensional theoretical model for determining susceptibility to the Unintentional Insider Threat in information systems security. All the objectives outlined for this study were achieved as discussed in Section 5.2 by developing a multi-dimensional model grounded in theory in addition to validating it through Structural Equation Modeling using data collected from a naturalistic field study. This thesis has articulated the process of developing this model in Chapter 2 and has justified the various decisions taken regarding theory and factor selection. Details of the design and execution of the naturalistic field study were outlined in Chapter 4. The validation of the model using Structural Equation Modeling was detailed in Chapter 5 and the findings discussed in Chapter 6.

The resulting model has been found to have a good explanatory power; performing better than models presented in recent studies. It is able to account for 28.7% of the variance in the Unintentional Insider Threat behavioural outcome. In addition, the model is able to account for 43.1% of the variance in Threat Detection, 19.1% of the variance in Threat Avoidance and 41.4% of the variance in Elaboration.

6.4 Implications of the Study

This study has made several contributions to theory, the body of knowledge, policy and practice, and these are discussed hereafter.

6.4.1 Theoretical contributions

This study has made bold theoretical contributions by integrating three theories and various factors unveiled in empirical studies in order to provide a unified multi-dimensional theoretical model for understanding the unintentional insider threat phenomenon. The model integrates the Elaboration Likelihood Model by Petty & Cacioppo (1986) and the Protection Motivation Theory by Rogers (1975, 1983). It also makes reference to the Technology Threat Avoidance Theory by Liang & Xue (2009, 2010).

The unified multi-dimensional theoretical model presents 22 independent variables and 1 dependent variable and 12 demographic variables. The causal relationships between the different variables are outlined and discussed in order to provide a better understanding and holistic representation of the unintentional insider threat. Many of the variables have not been tested together in previous studies and this provides new insights on their inter-relationships.

6.4.2 Contributions to the Body of Knowledge

In addition to presenting the theoretical model, this study has gone ahead to empirically validate this model. The data to validate the model was collected through a naturalistic field study that provides high ecological validity and allows its findings to be generalizable to wider contexts.

The model theorized and presented a total of 27 hypothesis which presented various causal relationships that would explain the unintentional insider threat phenomenon. Five of the hypotheses were not tested because their associated factors were dropped during the Exploratory Factor Analysis process. Twelve of the hypotheses were not supported by the data during the Structural Equation Modeling analysis; however, ten were supported and upheld.

Threat avoidance was not found to have a positive effect on the unintentional insider threat behaviour outcome but this effect was not found to be significant. The effect was theorized to be negative but the results showed it to be positive. This could be because insiders have an intention to avoid the threat but are not equipped and skilled enough to actually manifest the avoidance behaviour.

Three factors were examined in relation to the Coping Appraisal construct. These were: Response Efficacy, Response Cost and Perceived Benefit. Self-Efficacy was dropped during Exploratory Factor Analysis indicating the measurement items were not effective in measuring the factor. Response Efficacy was found to have a significant positive effect on Threat Avoidance. This means that insiders who have confidence on the use of available response measures are likely to avoid unintentional insider threats. Response Cost was found to have a significant negative effect on Threat Avoidance. This means that insiders who associate undesirable costs on response measures are less likely to avoid unintentional insider threats. Perceived Benefit was found to have a positive effect on Threat Avoidance but this effect was not significant. This means that insiders who associate benefits to taking response measures are more likely to avoid unintentional insider threats.

Three factors were proposed in relation to the organizational defenses and these were Policies, Technology and Security Education Training and Awareness (SETA). The Policies factor was dropped during Exploratory Factor Analysis. In addition, Technology and Security Education Training and Awareness factors were not found to have a significant effect on Threat Avoidance despite showing positive effects. The positive effect means that incorporation of these organizational defenses allows insiders to avoid unintentional insider threats.

Threat Detection was found to have a positive effect on Threat Avoidance but this effect was not found to be significant. The positive effect means that insiders who detect unintentional insider threats should also be motivated to avoid them.

Threat Detection was found to have a significant negative effect on the Unintentional Insider Threat Behaviour Outcome. This means that insiders who correctly perceive threats are also able to prevent themselves from falling victim.

Two factors were studied under the Threat Appraisal construct: Perceived Vulnerability and Perceived Severity. None of them were found to have a significant effect on Threat Detection. This could be because insiders who have a general awareness of their vulnerability and severity to UIT threats are not necessarily able to detect the threats when they actually present themselves.

Four factors were studied as part of the Knowledge construct. These were: Knowledge on Threat Domain, Detection Cues, Low Trust Determinants and High Trust Determinants. Knowledge on the Threat Domain showed a positive effect on Threat Detection, meaning insiders with a good understanding of UIT threats are likely to detect these threats. This effect was not significant. Knowledge on Detection Cues also showed a positive effect on Threat Detection and this effect was found to be significant. This means that insiders with a good understanding of Detection Cues are able to detect UIT threats. The Trust Determinants factor was initially theorized as a single factor. However, the Exploratory Factor Analysis revealed that there were two distinct dimensions to the factor. In fact the two dimensions had opposite effects on threat detection. Insiders who relied on Low Trust Determinants (that are easily manipulated during deception) are less likely to detect UIT threats. However, insiders who have knowledge on how to use High Trust Determinants (relating to credible security indicators) are more likely to detect UIT threats.

Elaboration is a factor representing the cognitive effort insiders expend in the evaluation of issue-relevant arguments of a deceptive scenario. Elaboration was found to have a positive and significant effect on Threat Detection. This means that insiders who consciously evaluate deceptive scenarios are more likely to detect UIT threats.

The effect of Elaboration was also examined with relation to Threat Avoidance. The results showed a positive and significant effect of Elaboration on Threat Avoidance. This means that insiders who consciously expend cognitive resources to evaluate deceptive scenarios are more likely to avoid UIT threats.

In addition, the effect of Elaboration on the Unintentional Insider Threat Behavioural Outcome was examined. Results showed that Elaboration has a negative effect on the behavioural outcome but this effect was not found to be significant. The negative effect means that insiders who consciously process deceptive scenarios are less likely to succumb to UIT threats.

Two factors were examined as part of the Attack Factors construct. These are: Quality of Argument and Persuasive Cues. The Attack Factors construct was included in the study in order to incorporate the attack factor element in the multi-dimensional model. Results showed that Quality of Argument has a positive and significant effect on Elaboration. This shows that deceptive scenarios with high quality of argument lead to high elaboration. Persuasive Cues were found to have a negative effect on Elaboration but this effect was not significant. The negative effect means that persuasive cues lead to low elaboration.

Similarly, two factors were examined as part of the Motivation to Process construct. These are: Involvement and Responsibility. Involvement was found to have a negative effect on Elaboration but this effect was not significant. The negative effect was contrary to the hypothesized positive effect. It could be because the measures were not expressed well enough to capture the correct direction of the effect. The Responsibility factor was found to have a positive and significant effect on Elaboration. This means that insiders who feel responsible with regards to issues presented in deceptive scenarios are more likely to consciously process those scenarios.

The three factors that were theorized as part of Ability to Process were examined as one latent factor due to the results of the Exploratory Factor Analysis. The Ability to Process was found to have a positive effect on Elaboration but this effect was not found to be significant. The positive effect means that insiders who have a high ability to process (because they are not distracted, emotionally manipulated or under pressure) are more likely to consciously process deceptive scenarios – thereby demonstrating high elaboration.

These contributions to the extant body of knowledge provide a better understanding of the Unintentional Insider Threat phenomenon. The contributions are particularly noteworthy because the effects of these factors have been examined in an integrated model and not in piecemeal as has been in previous studies.

6.4.3 Recommendations for Practice and Policy

The findings of this study provide insights that can positively transform how organizations address unintentional insider threats. The first recommendation is to address unintentional insider threats from multiple dimensions. This study has examined the factors relating to insiders, the organization and the attacker. As pointed out by Tetri & Vuorinen (2013), many organizations have tried to address the unintentional insider threat by focusing on the insider but have not paid as much attention on factors relating to organization defenses or the attacker's tactics. This has made many efforts ineffective. Future studies should build on this multi-dimensional approach in order to bring a more comprehensive and holistic understanding that can bring about effective solutions to the unintentional insider threats.

Second, the results of this study show that Threat Detection has a great effect on the Unintentional Insider Threat Behaviour Outcome. Organizations should therefore invest in measures that equip insiders with the ability to detect threats. The capability to detect threats can particularly be built through enhancing their knowledge on detection cues and trust determinants. Another factor that can particularly influence practice is Elaboration. The ability for insiders to consciously and objectively evaluate threat scenarios has a great impact on their threat detection and threat avoidance. This finding may not be clearly understood and addressed by current approaches to mitigate unintentional insider threats. Solutions should be built around countering attack factors that try to diminish insider's ability to examine deceptive scenarios. If insiders were to be repeatedly urged to pay attention to various threat scenarios that could target them, they would be less susceptible to unintentional insider threats.

Third, this research presents a credible naturalistic methodology that information security practitioners can use to assess organizational exposure to phishing threats. Such assessments can be done regularly and routinely compared against security metrics established by organizations based on their security goals. Assessment results can be tracked over a period of time and the effectiveness of implemented countermeasures examined to see their effectiveness in reducing insider susceptibility to attacks. Such naturalistic studies allow a direct observation of actual behaviour and this provides a more reliable assessment of susceptibility than thorough the use of self-reported questionnaires or surveys. It is important to obtain an ethical review and approval before conducting such assessments to ensure that the research protocol to be used protects participants from actual harm. In this study, approvals were obtained from the university's research office, ICT department and it's ethical Institutional Review Board (IRB). These layers of review protected the institution and its insiders from adverse effects during the staged phishing attack.

Fourth, this study presents instruments that organizations can use to assess their insider exposure to phishing attacks. The instruments are described in detail and the accompanying source code is provided in Appendix C, D and E. In this study, phishing was conducted using targeted spear phishing emails and also by setting up a phishing website. As demonstrated in this research, phishing instruments should be developed and designed with care to ensure they are convincing. Previous phishing attacks targeted at an organization should be studied and the characteristics of current attack methods established. In addition, pretext scenarios that are relevant to the current affairs at the organization should be chosen. These considerations during design match current attack tactics and also ensure that staged attacks are not easily dismissed without eliciting interaction from insiders who are susceptible and truly at risk.

Fifth, this study shows how malicious outsiders can target organizations by registering phishing domains that are deceptively similar to the organization's operational domain. The recommendation presented for practice and policy is that organizations should closely monitor domains that are very similar to their operational domains. These could be deceptive variants of their operational domain but also those ending with different suffixes such as .com, .org or even country suffices such as .or.ke, as was the case in this study. The information technology or security teams at the organization could probably go a step further to buy such domains instead of leaving them available for outsiders to acquire. It is not a very expensive venture since registering a domain could cost as low as 10 U.S. dollars per year, as was the case in this study. The organization need not buy all possible domains but those it considers very closely matched to its operational domain.

Sixth, this study also informs policy and practice by demonstrating how attackers can obtain a lot of information and system access using sophisticated embedded scripts. The phishing instruments used in this study were designed with active scripts that did a lot of background work and harvested system details. The background scripts are presented in Appendix E. This highlights a very important point that organizations and users need to appreciate. Phishing is not considered successful only when a person fills in sensitive information on a web form. Attackers collect valuable information right from the time a person opens their emails or loads their websites. Current phishing scams are very sophisticated and contain a lot of active scripts that harvest information from user systems and even install malware with minimal engagement and visibility by end users. Once an attacker is successful with even a few devices or accounts, they can work into the rest of the organization. Organizations should therefore provide additional layers of protection from such active content, for example, by disabling active scripts from being executed by email clients, web browsers and document processors. Users should also be made aware about current attack methods that make use of active content in emails, web pages, pdf documents and also, for example, through embedded macros in Microsoft documents. Attackers can succeed in security breaches through very little engagement with insiders and through seemingly harmless actions such as opening an email or loading a website.

Seventh, this research also informs policy and practice by demonstrating how vocal insiders can be a vital and effective countermeasure in curtailing active attacks. In this study, a prominent blogger was able to raise an alarm and rally action through social media. Within a few hours, an alert of the ongoing attack had been circulated throughout the institution. Organizations should invest in channels through which users can quickly report suspected attacks and through which information can be shared with the wider population so as to frustrate the efforts of attackers. They should turn each user on their system into an intrusion detection agent with the skill and capability to detect threats and sound an alarm for immediate action. Organizations should invest in staff who can monitor reporting channels including those from social media. Such staff can then engage information security incident response teams to quickly respond in the event of an ongoing attack and quickly contain the situation to limit damage to information systems.

6.5 Strengths of the Study

There are a number of strengths of this study. The first is that this study sought to establish a multi-dimensional theoretical foundation for the study of unintentional insider threats. A review of extant literature shows a deficiency in theoretically grounded studies on unintentional insider threats. In addition, the approaches used in previous studies have been piecemeal and have not examined multiple dimensions of the phenomenon.

Another strength is that this study has not only provided a theoretical model, it has gone ahead to validate the model empirically. The validation process used in this study has analyzed the integrated model using data from a naturalistic field study. The staging of a naturalistic field study that mimics real-world unintentional insider threats ensures that the data and findings from the study have high ecological validity. This is a key strength of this study.

In addition, data was not collected through observation of actual insider behaviour. The sole reliance of self-reported data could lead to inaccurate findings because insiders sometimes may not be aware of their unintentional weaknesses and may also shy away from honestly responding to embarrassing behaviour.

The use of the Structural Equation Modeling process for model validation is a key strength of this study. This is because the SEM process allows the fully integrated model to be analyzed in its entirety at the same time. Traditional statistical analysis techniques such as regression analysis examine models in piecemeal. Unlike these traditional approaches, the SEM process allows the interplay of relationships to be examined and to give integrated results.

6.6 Limitations of the Study

There are a couple of notable limitations of this study. Many of them relate to the staging of the naturalistic field experiment. Previous studies have pointed out that it is generally difficult to get research approvals to stage attacks in organizations for the purposes of research (Bakhshi et al., 2009; Finn & Jakobsson, 2007; Huber et al., 2009; Kumaraguru et al., 2008; Vishwanath et al., 2011; Wang et al., 2012). This was true for this study. A number of organizations refused to participate in the research and this delayed progress in the research.

In addition, when the naturalistic experiment was staged, it was discontinued after a short period of time leading to a shorter time for data collection than was intended. This was because the organization faced potential reputational damage when a prominent blogger posted an alert on a popular social media page calling for the organization to investigate a possible attack. The ICT team alerted the entire organization of the ongoing research and at that point requested the researcher to pull down the research website and to stop the staged experiment.

This led to the other limitation of this study. The amount of data collected for analysis was only 192 usable cases from a possible sample of 4,483 targeted insiders. This was far less than what had been hoped for. As highlighted by Wolf, Harrington, Clark, & Miller (2013) it is difficult to specify general sample size guidelines for structural equation modeling. Some guidelines have allowed minimum samples of 100 Anderson & Gerbing (1984) and others have given a ratio of 5 to 10 cases per parameter

Bentler & Chou (1987). It is important to note that the usable cases obtained were sufficient for the structural equation modeling because the model converged and in addition it adequately achieved overall fit criteria. However, the sample obtained impacts on the generalizability of the results of this study. Future studies should target larger sample sizes and with more diverse populations.

Finally, it should be pointed out that some of the factors (Self-Efficacy, Policies, Distraction, Emotions and Pressure) were dropped during exploratory and confirmatory factor analysis. In addition, some hypotheses were not supported by the data set during analysis. These affected the overall model prediction. Possible explanations for this are that the measurement items were not adequate and the factors were not well captured by this particular study. Future studies should examine how to adequately measure these factors and how to improve overall model prediction.

6.7 Suggestions for Further Research

Lessons learnt from the undertaking of this study can give meaningful recommendations that can successfully guide future research. First, this study presents a unified multi-dimensional model that can provide a good starting foundation for future studies. The predictive power of the model can still be further improved, particularly by including additional factors or examining relationships in a different configuration. There could be multiple path models that can be examined to see which ones give a better goodness-of-fit and predictive power.

Second, the proposed multi-dimensional model can be tested using data from different organizations, diverse populations and contexts in order to allow a wider generalization of findings. This would also provide a larger dataset that would give a better validation of such a complex integrated model.

Third, some lessons can be learnt to help future studies effectively identify users susceptible to unintentional insider threats. It would be important to confirm that user accounts targeted in future studies are operational and that no delivery failures take place. Additionally, it is important to establish whether users regularly engage with their emails or if they interact with other communication channels such as social media or chat groups. In addition, increasing the study period can also give users a longer time to review staged attacks and thereby increase the effectiveness of identifying those truly susceptible.

Fourth, it is recommended that future research undertake longitudinal studies that can examine the unintentional insider threat phenomenon over longer period of time and using multiple threat scenarios. This study used one scenario based on social engineering through phishing. It would be important to test the model using data from different threat scenarios.

REFERENCES

Aldawood, H., & Skinner, G. (2018). *Educating and Raising Awareness on Cyber Security Social Engineering: A Literature Review*. Presented at the International Conference on Teaching, Assessment, and Learning for Engineering (TALE), Wollongong, NSW, Australia.

Algarni, A. (2019). What Message Characteristics Make Social Engineering Successful on Facebook: The Role of Central Route, Peripheral Route, and Perceived Risk. *Information*, *10*(211). https://doi.org/10.3390/info10060211

Ali, A. (2015). *Social Engineering: Phishing latest and future techniques*. Retrieved from https://www.researchgate.net/publication/274194484

Allen, M. (2006). *Social Engineering: A Means To Violate A Computer System*. SANS Institute.

Alsharnouby, M., Alaca, F., & Chiasson, S. (2015). Why phishing still works: User strategies for combating phishing attacks. *International Journal of Human-Computer Studies, 82*, 69–82. http://dx.doi.org/10.1016/j.ijhcs.2015.05.005

Anandpara, V., Dingman, A., Jakobsson, M., Liu, D., & Roinestad, H. (2007). *Phishing IQ Tests Measure Fear, Not Ability*. 362–366. Trinidad and Tobago: Springer Berlin Heidelberg.

Andersen, D., Cappelli, D. M., Gonzalez, J. J., Mojtahedzadeh, M., Moore, A. P., Rich, E., … Zagonel, A. (2004, February). *Preliminary System Dynamics Maps of the Insider Cyber-threat Problem*. Retrieved from www.cert.org/archive/pdf/InsiderThreatSystemDynamics.pdf

Anderson, C. L., & Agarwal, R. (2010). Practicing Safe Computing: A Multimethod Empirical Examination of Home Computer User Security Behavioral Intentions. *MIS Quarterly*, *34*(3), 613–643.

Anderson, J. C., & Gerbing, D. W. (1984). The effect of sampling error on convergence, improper solutions, and goodness-of-fit indices for maximum likelihood confirmatory factor analysis. *Psychometrika*, *49*(2), 155–173.

Angst, C. M., & Agarwal, R. (2009). Adoption of Electronic Health Records in the Presence of Privacy Concerns: The Elaboration Likelihood Model and Individual Persuasion. *MIS Quarterly*, *33*(2), 339–370.

Applegate, S. D. (2009). Social Engineering: Hacking the Wetware! *Information Security Journal: A Global Perspective*, *18*, 40–46. https://doi.org/10.1080/19393550802623214

APWG, A.-P. W. G. (2016). *Phishing Activity Trends Report: 1st Quarter 2016*. Retrieved from https://docs.apwg.org/reports/apwg_trends_report_q1_2016.pdf

APWG, A.-P. W. G. (2017). *Phishing Activity Trends Report: 4th Quarter 2016*. Retrieved from https://docs.apwg.org/reports/apwg_trends_report_q4_2016.pdf

APWG, A.-P. W. G. (2018). *Phishing Activity Trends Report: 3rd Quarter 2017*. Retrieved from http://docs.apwg.org/reports/apwg_trends_report_q3_2017.pdf

Arachchilage, N. A. G., & Love, S. (2013). A Game Design Framework for Avoiding Phishing Attacks. *Computers in Human Behavior*, *29*, 706–714.

Ashford, W. (2019a, May 30). Hackers targeting UK universities a threat to national security. *Computer Weekly*. Retrieved from https://www.computerweekly.com/news/252464169/Hackers-targeting-UK-universities-a-threat-to-national-security

Ashford, W. (2019b, July 23). Phishing attack highlights cyber security need at universities. *Computer Weekly*. Retrieved from https://www.computerweekly.com/news/252467214/Phishing-attack-highlights-cyber-security-need-at-universities

Awang, Z. (2012). *A Handbook on SEM (Structural Equation Modeling) Using AMOS Graphic* (2nd ed.). Kota Baharu: Universiti Teknologi Mara Kelantan.

Awang, Z. (2015). *SEM Made Simple: A Gentle Approach to Learning SEM*. MPWS.

Aytes, K., & Connolly, T. (2004). Computer Security and Risky Computing Practices: A Rational Choice Perspective. *Journal of Organizational and End User Computing*, *16*(3), 22–40.

Aytes, K., & Conolly, T. (2003). A Research Model for Investigating Human Behavior Related to Computer Security. *Proceedings of the 2003 American Conference On Information Systems (AMCIS)*. Presented at the Tampa, Florida. Tampa, Florida.

Bakhshi, T., Papadaki, M., & Furnell, S. (2009). Social Engineering: Assessing Vulnerabilities in Practice. *Information Management & Computer Security*, *17*(1), 53–63. https://doi.org/10.1108/09685220910944768

Bandura, A. (1977). Self-Efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, *84*(2), 191–215.

Barth, B. (2016, August 1). Don’t be like “Mike”: Authorities arrest mastermind of $60M online scam operation. *SC Magazine*.

Baruch, Y., & Holtom, B. C. (2008). Survey Response Rate Levels and Trends In Organizational Research. *Human Relations*, *61*(8), 1139–1160. https://doi.org/10.1177/0018726708094863

BBC News. (2016, August 1). Online fraud: Top Nigerian scammer arrested. *BBC News*. Retrieved from http://www.bbc.com/news/world-africa-36939751

Bentler, P. M., & Chou, C.-P. (1987). Practical Issues in Structural Modeling. *Sociological Methods & Research*, *16*(1), 78–117.

Bezuidenhout, M., Mouton, F., & Venter, H. S. (2010). Social Engineering Attack Detection Model: SEADM. *Information Security for South Africa (ISSA)*. IEEE.

Bhattacherjee, A. (2012). *Social Science Research: Principles, Methods, and Practices* (2nd ed.). University of South Florida Scholar Commons.

Bhattacherjee, A., & Sanford, C. (2006). Influence Processes for Information Technology Acceptance: An Elaboration Likelihood Model. *MIS Quarterly*, *30*(4), 805–825.

Bishop, M., & Gates, C. (2008). Defining the Insider Threat. *The 4th Annual Workshop: Developing Strategies to Meet the Cyber Security and Information Intelligence Challenges Ahead*. Presented at the Cyber Security and Information Intelligence Research Workshop (CSIIRW), Oak Ridge, Tennessee USA.

Bojmaeh, H. Y. (2015). The Main Factors Influencing Information Security Behavior. *International Journal of Science and Engineering Applications*, *4*(6). https://doi.org/10.7753/IJSEA0406.1004

Boone, H. N., & Boone, D. A. (2012). Analyzing Likert Data. *Journal of Extension*, *50*(2), 1–5.

Bose, I., & Leung, A. C. M. (2007). Unveiling the Mask of Phishing: Threats, Preventive Measures, and Responsibilities. *Communications Of The Association For Information Systems*, *19*(1), 544–566.

Boss, S. R., Kirsch, L. J., Angermeier, I., Shingler, R. A., & Boss, R. W. (2009). If someone is watching, I'll do what I'm asked: Mandatoriness, control, and information security. *European Journal of Information Systems*, *18*(2), 151–164.

Bowen, B. M., Salem, M. B., Hershkop, S., Keromytis, A. D., & Stolfo, S. J. (2009). Designing host and network sensors to mitigate the insider threat. *Security & Privacy*, *7*(6), 22–29.

Brehm, J. W. (1966). *A Theory of Psychological Reactance*. New York: Academic Press.

Brewer, M. B., & Crano, W. D. (2014). Research Design and Issues of Validity. In H. T. Reis & C. M. Judd (Eds.), *Handbook of Research Methods in Social and Personality Psychology* (2nd ed., pp. 3–16). New York: Cambridge University Press.

Broadhurst, R., Skinner, K., Sifniotis, N., Matamoros-Macias, B., & Ipsen, Y. (2019). Phishing and Cybercrime Risks in a University Student Community. *International Journal of Cybersecurity Intelligence & Cybercrime*, *2*(1), 4–23.

Brown, T. A. (2006). *Confirmatory Factor Analysis for Applied Research*. New York, London: Guilford Press.

Buckley, O., Nurse, J. R., Legg, P. A., Goldsmith, M., & Sadie, C. (2014). *Reflecting on the ability of enterprise security policy to address accidental insider threat*. 8–15. IEEE.

Bulgurcu, B., Cavusoglu, H., & Benbasat, I. (2010). Information Security Policy Compliance: An Empirical Study of Rationality-Based Beliefs and Information Security Awareness. *MIS Quarterly*, *34*(3), 523–548.

Buller, D. B., & Burgoon, J. K. (1996). Interpersonal Deception Theory. *Communication Theory*, *6*(3), 203–242.

Burstein, A. J. (2008). *Toward a Culture of Cybersecurity Research* (UC Berkeley Public Law Research Paper No. 1113014). Retrieved from http://dx.doi.org/10.2139/ssrn.1113014

Butavicius, M., Parsons, K., Pattinson, M., & McCormac, A. (2015). *Breaching the Human Firewall: Social engineering in Phishing and Spear-Phishing Emails*. Presented at the Australasian Conference on Information Systems. Retrieved from https://arxiv.org/ftp/arxiv/papers/1606/1606.00887.pdf

Butavicius, M., Parsons, K., Pattinson, M., McCormac, A., Calic, D., & Lillie, M. (2017). Understanding Susceptibility to Phishing Emails: Assessing the Impact of Individual Differences and Culture. *Proceedings of the Eleventh International Symposium On*, 12–23.

CA Technologies. (2018). *Insider Threat Report*. Retrieved from https://www.ca.com/content/dam/ca/us/files/ebook/insider-threat-report.pdf

Canfield, C. I., Fischhoff, B., & Davis, A. (2016). Quantifying Phishing Susceptibility for Detection and Behavior Decisions. *Human Factors*, *58*(8), 1158–1172. https://doi.org/10.1177/0018720816665025

Cappelli, D., Moore, A., & Trzeciak, R. (2012). *The CERT Guide to Insider Threats: How to Prevent, Detect, and Respond to Information Technology Crimes*. Addison-Wesley.

Carnegie Mellon University. (2013). *U.S. State of Cybercrime Survey*. Software Engineering Institute CERT Program.

Carver, C. S. (2006). Approach, Avoidance and the Self-Regulation of Affect and Action. *Motivation and Emotion*, *30*, 105–110.

Carver, C. S., & Scheier, M. F. (1982). Control Theory: A Useful Conceptual Framework for Personality-Social, Clinical and Health Psychology. *Psychological Bulletin*, *92*(1), 111–135.

Castelfranchi, C., & Falcone, R. (2010). *Trust Theory: A Socio-Cognitive and Computational Model*. John Wiley & Sons Ltd.

Cataldo, E. F., Johnson, R. M., Kellstedt, L. A., & Milbrath, L. W. (1970). Card Sorting as a Technique for Survey Interviewing. *The Public Opinion Quarterly*, *34*(2), 202–215.

CERT, I. T. T. (2013). *Unintentional Insider Threats: A Foundational Study*. Retrieved from http://resources.sei.cmu.edu/asset_files/TechnicalNote/2013_004_001_58748.pdf

Chaiken, S. (1980). Heuristic Versus Systematic Information Processing and the Use of Source Versus Message Cues in Persuasion. *Journal of Personality and Social Psychology*, *39*(5), 752–766.

Chinchani, R., Iyer, A., Ngo, H. Q., & Upadhyaya, S. (2005). *Towards A Theory Of Insider Threat Assessment*. 108–117. IEEE.

Chuenchujit, T. (2016). *A Taxonomy of Phishing Research* (Master of Science in Computer Science). University of Illinois at Urbana-Champaign, Urbana, Illinois.

Cialdini, R. B. (2001). *Influence: Science and Practice* (4th ed.). Boston MA: Allyn & Bacon.

Cimpanu, C. (2016, April 28). Anonymous Hackers Leak 1 TB of Documents from Kenya's Ministry of Foreign Affairs. Retrieved September 12, 2016, from Softpedia website: http://news.softpedia.com/news/anonymous-hackers-leak-1tb-of-documents-from-kenya-s-ministry-of-foreign-affairs-503518.shtml

Cochran, W. G. (1977). *Sampling Techniques* (3rd ed.). New York: Wiley.

Collins, M., Theis, M., Trzeciak, R., Strozer, J., Clark, J., Costa, D., … Moore, A. (2016). *Common Sense Guide to Mitigating Insider Threats* (5th ed.). Carnegie Mellon University.

Creswell, J. W. (2003). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches* (2nd ed.). SAGE Publications.

Cummings, A., Lewellen, T., McIntire, D., Moore, A. P., & Trzeciak, R. (2012). *Insider Threat Study: Illicit Cyber Activity Involving Fraud in the U.S. Financial Services Sector*. Software Engineering Institute, Carnegie Mellon University.

Curran, P. J., West, S. G., & Finch, J. F. (1996). The Robustness of Test Statistics to Nonnormality and Specification Error in Confirmatory Factor Analysis. *Psychological Methods*, *1*(1), 16–29.

Cyveillance. (2015). *The Cost of Phishing: Understanding the True Cost Dynamics Behind Phishing Attacks*. Cyveillance.

D'Arcy, J., Hovav, A., & Galletta, D. (2009). User Awareness of Security Countermeasures and Its Impact on Information Systems Misuse: A Deterrence Approach. *Information Systems Research*, *20*(1), 79–98.

Dhamija, R., & Tygar, J. D. (2005). The battle against phishing: Dynamic security skins. *Proceedings of the 2005 Symposium on Usable Privacy and Security*, 77–88. ACM.

Dhamija, R., Tygar, J. D., & Hearst, M. (2006). Why Phishing Works. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 581–590. ACM.

Diener, E., & Crandall, R. (1978). *Ethics in Social and Behavioral Research*. University of Chicago Press.

Dodge, R. C., Carver, C., & Ferguson, A. J. (2007). Phishing for User Security Awareness. *Computers & Security*, *26*, 73–80.

Downs, J. S., Holbrook, M., & Cranor, L. F. (2006). Decision Strategies and Susceptibility to Phishing. *Proceedings of the Second Symposium on Usable Privacy and Security*, 79–90. ACM.

Downs, J. S., Holbrook, M., & Cranor, L. F. (2007). Behavioral Response to Phishing Risk. *Proceedings of the Anti-Phishing Working Groups 2nd Annual ECrime Researchers Summit*, 37–44. ACM.

Eagly, A. H., & Chaiken, S. (1993). *The Psychology of Attitudes*. Harcourt Brace Jovanovich College Publishers.

Egelman, S., Cranor, L. F., & Hong, J. (2008). You've been warned: An empirical study of the effectiveness of web browser phishing warnings. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM.

Festinger, L. (1957). Cognitive Dissonance Theory. In *Primary Prevention of HIV/AIDS: Psychological Approaches*. Newbury Park, California: SAGE Publications.

Fette, I., Sadeh, N., & Tomasic, A. (2007). Learning to Detect Phishing Emails. *Proceedings of the 16th International Conference on World Wide Web*, 649–656. ACM.

Field, A. (2009). *Discovering Statistics Using SPSS* (3rd ed.). SAGE Publications Ltd.

Finn, P., & Jakobsson, M. (2007). Designing and Conducting Phishing Experiments. *IEEE Technology and Society Magazine, Special Issue on Usability and Security*, *26*(1), 46–58.

Fire Eye. (2015, April). *APT30 And The Mechanics Of A Long-Running Cyber Espionage Operation*. Retrieved from https://www.fireeye.com/current-threats/threat-intelligence-reports.html

Fire Eye. (2017, January). *APT28: At The Center Of The Storm*. Retrieved from https://www.fireeye.com/current-threats/threat-intelligence-reports.html

Fitzgerald, B., & Howcroft, D. (1998). Towards Dissolution of the IS Research Debate: From Polarization to Polarity. *Journal of Information Technology*, *13*(4), 313–326.

Flynn, L., Huth, C., Trzeciak, R., & Buttles, P. (2013). *Best Practices Against Insider Threats in All Nations*. Retrieved from http://resources.sei.cmu.edu/asset_files/TechnicalNote/2013_004_001_59084.pdf

Fogg, B. J., Marshall, J., Laraki, O., Osipovich, A., Varma, C., Fang, N., & Jyoti, P. (2001). What makes Web sites credible? A report on a large quantitative study. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 61–68. ACM.

Fornell, C., & Larcker, D. F. (1981). Structural Equation Models with Unobservable Variables and Measurement Error: Algebra and Statistics. *Journal of Marketing Research*, 382–388.

Friedman, B., Hurley, D., Howe, D. C., Felten, E., & Nissenbaum, H. (2002). Users' Conceptions of Web Security: A Comparative Study. *Extended Abstracts on Human Factors in Computing Systems*, 746–747. ACM.

Garera, S., Provos, N., Chew, M., & Rubin, A. D. (2007). A Framework for Detection and Measurement of Phishing Attacks. *Proceedings of the 2007 ACM Workshop on Recurring Malcode*. ACM.

Gaskin, J. (2012). Ten Steps for Formulating a Decent Quantitative Model. Retrieved August 18, 2014, from Gaskination's StatWiki website: http://statwiki.kolobkreations.com/

Gaskin, J. (2019). Guidelines. Retrieved July 30, 2019, from Gaskination's StatWiki website: http://statwiki.kolobkreations.com

Goel, S., Williams, K., & Dincelli, E. (2017). Got Phished? Internet Security and Human Vulnerability. *Journal of the Association for Information Systems*, *18*(1), 22–44.

Government of Kenya. (2012, October). *GoK Cybersecurity Masterplan Draft*.

Grazioli, S. (2004). Where Did They Go Wrong? An Analysis of the Failure of Knowledgeable Internet Consumers to Detect Deception Over the Internet. *Group Decision and Negotiation*, *13*(2), 149–172.

Grazioli, S., & Jarvenpaa, S. (2001). Tactics Used Against Consumers as Victims of Internet Deception. *AMCIS 2001 Proceedings*, 810–815. Boston.

Greener, S. (2008). *Business Research Methods*. BookBoon.

Greitzer, F. L., Strozer, J., Cohen, S., Bergey, J., Cowley, J., Moore, A., & Mundie, D. (2014). Unintentional Insider Threat: Contributing Factors, Observables, and Mitigation Strategies. *47th Hawaii International Conference on System Science*. IEEE.

Gupta, B. B., Arachchilage, N. A. G., & Psannis, K. E. (2018). Defending against phishing attacks: Taxonomy of methods, current issues and future directions. *Telecommunication Systems*, *67*(2), 247–267.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2009). *Multivariate Data Analysis* (7th ed.).

Hair Jr., J. F., Matthews, L. M., Matthews, R. L., & Sarstedt, M. (2017). PLS-SEM or CB-SEM: Updated Guidelines on which Method to Use. *International Journal of Multivariate Data Analysis*, *1*(2), 107–123.

Hassan, S. A., & Abu Bakar, K. (2009). Exploratory Factor Analysis versus Confirmatory Factor Analysis. In F. Abd Rahman, F. P. Shafiee, & H. Elias (Eds.), *Teachers' Learning, Curiculum Inovations and Knowledge Applications*.

Herath, T., & Rao, R. H. (2009a). Encouraging information security behaviors in organizations: Role of penalties, pressures and perceived effectiveness. *Decision Support Systems*, *47*(2), 154–165.

Herath, T., & Rao, R. H. (2009b). Protection motivation and deterrence: A framework for security policy compliance in organisations. *European Journal of Information Systems*, *18*(2), 106–125.

Hernandez, E., Regalado, D., & Villeneuve, N. (2015, July). *An Insider Look Into the World of Nigenrian Scammers*. Retrieved from https://www.fireeye.com/current-threats/threat-intelligence-reports.html

Homoliak, I., Toffalini, F., Guarnizo, J., & Elovici, Y. (2018). Insight into Insiders and IT: A Survey of Insider Threat Taxonomies, Analysis, Modeling, and Countermeasures. *ACM Computing Surveys*, *99*(99). Retrieved from https://arxiv.org/pdf/1805.01612.pdf

Huber, M., Kowalski, S., Nohlberg, M., & Tjoa, S. (2009). *Towards Automating Social Engineering Using Social Networking Sites*. *3*, 117–124. IEEE.

Hunt, M. K., Hopko, D. R., Bare, R., Lejuez, C. W., & Robinson, E. V. (2005). Construct Validity of the Balloon Analog Risk Task (BART) Associations with Psychopathy and Impulsivity. *Assessment*, *12*(4), 416–428.

ISACA. (2012). *COBIT 5 For Information Security*. Retrieved from www.isaca.org/cobit5info-sec

ISO. (2013). The ISO 27000 Directory. Retrieved March 22, 2013, from http://www.27000.org/index.htm

Jagatic, T., Johnson, N., Jakobsson, M., & Menczer, F. (2007). Social Phishing. *Communications of the ACM*, *50*(10), 94–100.

Jakobsson, M. (2005). Modeling and Preventing Phishing Attacks. *Financial Cryptography*, *5*.

Jakobsson, M., & Myers, S. (2006). *Phishing and Counter-Measures: Understanding the Increasing Problem of Electronic Identity Theft*. Wiley-Interscience.

Jakobsson, M., & Ratkiewicz, J. (2006). Designing Ethical Phishing Experiments: A study of (ROT13) rOnl query features. *Proceedings of the 15th International Conference on World Wide Web*. ACM.

Jakobsson, M., Tsow, A., Shah, A., Blevis, E., & Lim, Y.-K. (2007). What Instills Trust? A Qualitative Study of Phishing. *Financial Cryptography and Data Security*, 356–361. Springer Berlin Heidelberg.

James, L. (2005). *Phishing Exposed*. Syngress.

Johnson, P. E., Grazioli, S., Jamal, K., & Berryman, R. G. (2001). Detecting Deception: Adversarial Problem Solving in a Low Base-Rate World. *Cognitive Science*, *25*(3), 355–392.

Johnson, P. E., Grazioli, S., Jamal, K., & Zualkernan, I. A. (1992). Success and Failure in Expert Reasoning. *Organizational Behavior and Human Decision Processes*, *53*(2), 173–203.

Johnston, A. C., & Warkentin, M. (2010). Fear Appeals and Information Security Behaviors: An Empirical Study. *MIS Quarterly*, *34*(3), 549–566.

Jones, H. S., & Towse, J. (2018). Examinations of Email Fraud Susceptibility: Perspectives From Academic Research and Industry Practice. In *Psychological and Behavioral Examinations in Cyber Security* (pp. 80–97). IGI Global.

Kandias, M., Mylonas, A., Virvilis, N., Theoharidou, M., & Gritzalis, D. (2010). An Insider Threat Prediction Model. In *Trust, Privacy and Security in Digital Business* (pp. 26–37). Springer Berlin Heidelberg.

Karakasiliotis, Furnell, S. M., & Papadaki, M. (2006). Assessing End-User Awareness of Social Engineering and Phishing. *Proceedings of 7th Australian Information Warfare and Security Conference*. Presented at the Perth Western Australia. Perth Western Australia: Edith Cowan University.

Kleitman, S., Law, M. K. H., & Kay, J. (2018). It’s the deceiver and the receiver: Individual differences in phishing susceptibility and false positives with item profiling. *PLOS ONE*, *13*(10), e0205089. https://doi.org/10.1371/journal.pone.0205089

Kline, R. B. (2013). Exploratory and Confirmatory Factor Analysis. In Y. Petscher & C. Schatsschneider (Eds.), *Applied Quantitative Analysis in the Social Sciences* (pp. 171–207). New York: Routledge.

Kumaraguru, P., Cranshaw, J., Acquisti, A., Cranor, L., Hong, J., Blair, M. A., & Pham, T. (2009). School of Phish: A Real-Word Evaluation of Anti-Phishing Training. *Proceedings of the 5th Symposium on Usable Privacy and Security (SOUPS)*. Presented at the Mountain View, CA, USA. Mountain View, CA, USA: ACM.

Kumaraguru, P., Rhee, Y., Acquisti, A., Cranor, L. F., Hong, J., & Nunge, E. (2007). Protecting People from Phishing: The Design and Evaluation of an Embedded Training Email System. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 905–914. ACM.

Kumaraguru, P., Rhee, Y., Sheng, S., Hasan, S., Acquisti, A., Cranor, L. F., & Hong, J. (2007). Getting Users to Pay Attention to Anti-Phishing Education: Evaluation of Retention and Transfer. *Proceedings of the Anti-Phishing Working Groups 2nd Annual ECrime Researchers Summit*, 70–81. ACM.

Kumaraguru, P., Sheng, S., Acquisti, A., Cranor, L. F., & Hong, J. (2007). Teaching Johnny Not to Fall for Phish. *CM Transactions on Internet Technology (TOIT)*, *10*(2), 7.

Kumaraguru, P., Sheng, S., Acquisti, A., Cranor, L., & Hong, J. (2008). *Lessons From a Real World Evaluation of Anti-Phishing Training*. Presented at the eCrime Researcher's Summit.

LaRose, R., Rifon, N. J., & Enbody, R. (2008). Promoting Personal Responsibility for Internet Safety. *Communications of the ACM*, *51*(3), 71–76.

Lee, Y., & Kozar, K. A. (2005). Investigating Factors Affecting the Adoption of Anti-spyware Systems. *Communications of the ACM*, *48*(8), 72–77.

Lee, Y., & Larsen, K. R. (2009). Threat or Coping Appraisal: Determinants of SMB Executives' Decision to Adopt Anti-malware Software. *European Journal of Information Systems*, *18*, 177–187.

Lejuez, C. W., Read, J. P., Kahler, C. W., Richards, J. B., Ramsey, S. E., Stuart, G. L., … Brown, R. A. (2002). Evaluation of a Behavioral Measure of Risk Taking: The Balloon Analogue Risk Task (BART). *Journal of Experimental Psychology: Applied*, *8*(2), 75–84.

Levine, T. R. (2014). Truth-default Theory: A Theory of Human Deception and Deception Detection. *Journal of Language and Social Psychology*, *33*(4), 378–392. https://doi.org/10.1177/0261927X14535916

Liang, H., & Xue, Y. (2009). Avoidance of Information Technology Threats: A Theoretical Perspective. *MIS Quarterly*, *33*(1), 71–90.

Liang, H., & Xue, Y. (2010). Understanding Security Behaviors in Personal Computer Usage: A Threat Avoidance Perspective. *Journal of the Association for Information Systems (JAIS)*, *11*(7), 394–413.

Likert, R. (1932). A Technique for the Measurement of Attitudes. *Archives of Psychology*, *22*, 1–55.

Luo, X. (Robert), Brody, R., Seazzu, A., & Burd, S. (2011). Social Engineering: The Neglected Human Factor for Information Security Management. *Information Resources Management Journal (IRMJ)*, *24*(3), 1–8.

Luo, X. (Robert), Zhang, W., Burd, S., & Seazzu, A. (2013). Investigating Phishing Victimization with the Heuristic-Systematic Model: A Theoretical Framework and an Exploration. *Computers & Security*, *38*, 28–38.

Mandiant. (2004). *APT1: Exposing One of China's Cyber Espionage Unitsâ€ " FireEye*. Retrieved from https://www.fireeye.com/content/dam/fireeye-www/services/pdfs/mandiant-apt1-report.pdf

Mandiant. (2010). *M-Trends: The Advanced Persistent Threat*. Retrieved from https://www.fireeye.com/blog/threat-research/2010/01/m-trends-advanced-persistent-threat-malware.html

Marsh, H. W., Morin, A. J. S., Parker, P. D., & Kaur, G. (2014). Exploratory Structural Equation Modeling: An Integration of the Best Features of Exploratory and Confirmatory Factor Analysis. *Annual Review of Clinical Psychology*, *10*, 85–110.

Martinez-Moyano, I. J., Conrad, S. H., & Andersen, D. F. (2011). Modeling behavioral considerations related to information security. *Computers & Security*, *30*(6–7), 397–409.

Mera, A. (2015). *Unintentional Insider Threat: Policy, Training and Technologies to Mitigate End User Risk*. Retrieved from https://alejandromera.com/blog/wp-content/uploads/2015/06/UnintentionInsider.pdf

Metzger, M. J. (2007). Making Sense of Credibility on the Web: Models for Evaluating Online Information and Recommendations for Future Research. *Journal of the American Society for Information Science and Technology*, *58*(13), 2078–2091.

Miller, R. C., & Wu, M. (2005). Fighting Phishing at the User Interface. In L. Cranor & S. L. Garfinkel (Eds.), *Security and Usability: Designing Secure Systems that People Can Use*. O'Reilly.

Ministry of Information Communications and Technology. (2014, February). *National Cybersecurity Strategy*.

Mitnick, K. D., & Simon, W. L. (2002). *The Art of Deception: Controlling the Human Element of Security*. John Wiley & Sons.

Mohebzada, J. G., El Zarka, A., Bhojani, A. H., & Darwish, A. (2012). *Phishing in a University Community: Two Large Scale Phishing Experiments*. 249–254. IEEE.

Mumo, M. (2012, July). EA banks lost US$48.3 million to fraud-Deloitte survey. *Business Daily*. Retrieved from http://www.businessdailyafrica.com/Corporate-News/-/539550/1467718/-/xhlmfs/-/index.html

Myyry, L., Siponen, M., Pahnila, S., Vartiainen, T., & Anthony, V. (2009). What levels of moral reasoning and values explain adherence to information security rules? An empirical study. *European Journal of Information Systems*, *18*(2), 126–139.

Nieles, M., Dempsey, K., & Pillitteri, V. Y. (2017). *An Introduction to Information Security* (No. NIST.SP.800-12r1; p. 91). Retrieved from National Institute of Standards and Technology website: https://doi.org/10.6028/NIST.SP.800-12r1

Norusis, M. J. (2012). Chapter 13: Cluster Analysis. In *IBM SPSS Statistics 19 Statistical Procedures Companion* (pp. 375–404). Prentice Hall.

Nunnally, J., & Bernstein, I. (1994). *Psychometric Theory* (2nd ed.). New York: McGraw Hill.

Nyayieka, I. (2019a, January 2). Telecoms regulator warns over harmful software. Retrieved January 17, 2019, from Business Daily website: https://www.businessdailyafrica.com/corporate/tech/Telecoms-regulator-warns-over-harmful-software/4258474-4918250-14xtryjz/index.html

Nyayieka, I. (2019b, January 9). Regulator warns of rise in cybersecurity threats. Retrieved January 17, 2019, from Business Daily website: https://www.businessdailyafrica.com/corporate/tech/Regulator-warns-of-rise-in-cybersecurity-threats/4258474-4927994-c4k826z/index.html

Obulutsa, G. (2016, April 28). Hackers leak stolen Kenyan foreign ministry documents. Retrieved September 12, 2016, from Reuters website: http://www.reuters.com/article/us-cyber-kenya-idUSKCN0XP2K5

Obura, F. (2018, December 19). Cases of cyber attack in Kenya rise to 3.8 million. Retrieved January 17, 2019, from The Standard website:

https://www.standardmedia.co.ke/article/2001306810/cases-of-cyber-attack-in-kenya-rise-to-3-8-million

Odhiambo, H. (2019, January 3). Communications Authority warns organizations against a data stealing malware. Retrieved January 17, 2019, from CIO East Africa website: https://www.cio.co.ke/communications-authority-warns-organizations-against-a-data-stealing-malware/

Ophoff, J., Jensen, A., Sanderson-Smith, J., Porter, M., & Johnston, K. (2014). A Descriptive Literature Review and Classification of Insider Threat Research. *Proceedings of Informing Science & IT Education Conference (InSITE)*, 221–223.

Osongo, D. (2019, January 1). Communications Authority issues computer virus alert. Retrieved January 17, 2019, from The Standard website: https://www.standardmedia.co.ke/article/2001308011/communications-authority-issues-computer-virus-alert

Pahnila, S., Siponen, M., & Mahmood, A. (2007). Employees' Behavior towards IS Security Policy Compliance. *Proceedings of the 40th*. Presented at the Hawaii International Conference on System Sciences, Hawaii.

Parsons, M. H. (1974). What happened at Hawthorne? *Science*, *183*(4128), 922–932.

Peltier, T. R. (2006). Social Engineering: Concepts and Solutions. *Information Systems Security*, *15*(5), 13–21.

Petty, R. E. (1994). Two Routes to Persuasion: State of the Art. In G. d'Ydewalle, P. Eelen, & P. Berteleson (Eds.), *International Perspectives on Psychological Science* (Vol. 2, pp. 229–247). Hillsdale, NJ: Erlbaum.

Petty, R. E., & Cacioppo, J. T. (1986). The Elaboration Likelihood Model of Persuasion. *Advances in Experimental Social Psychology*, *19*.

Petty, R. E., & Wegener, D. T. (1998). Attitude Change: Multiple Roles for Persuasion Variables. In *Handbook of Social Psychology* (4th ed., pp. 323–390). New York: McGraw Hill.

Petty, R. E., & Wegener, D. T. (1999). The Elaboration Likelihood Model: Current Status and Controversies. In S. Chaiken & Y. Trope (Eds.), *Dual-Process Theories in Social Psychology* (pp. 37–72). New York, NY, US: Guilford Press.

Pfleeger, S. L., & Stolfo, S. J. (2009). Addressing the Insider Threat. *Security & Privacy*, *7*(6), 10–13.

PhishTank. (2016). PhishTank Stats. Retrieved October 14, 2016, from PhishTankâ€ " Statisitics about phishing activity and PhishTank Usage website: https://www.phishtank.com/stats.php

Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common Method Biases in Behavioral Research: A Critical Review of the Literature and Recommended Remedies. *Journal of Applied Psychology*, *88*(5), 879–903.

proofpoint. (2019a). *Protecting People: A Quarterly Analysis of Highly Targeted Cyber Attacks*. Retrieved from https://www.proofpoint.com/us/resources/threat-reports/quarterly-threat-analysis

proofpoint. (2019b). *State of the Phish Report*. Retrieved from https://www.proofpoint.com/state-of-the-phish

PwC. (2018). *2018 Global Economic Crime Survey: Kenya Report* (p. 28). Retrieved from https://www.pwc.com/ke/en/assets/pdf/gecs-2018-report.pdf

PWC, CERT, C. M. U., USSS, U. S. S., & CSO, M. (2013). *U.S. State of Cybercrime Survey*.

PWC, CSO, M., CERT, C. M. U., & USSS, U. S. S. (2015). *Key Findings from the 2015 U.S. State of Cybercrime Survey*. Retrieved from https://www.pwc.com/us/en/services/consulting/library/us-cybercrime-survey-2015.html

Raulot, A. (2019). *Bypassing phishing protections with email authentication* (Master of Security and Network Engineering, University of Amsterdam). Retrieved from https://delaat.net/rp/2018-2019/p61/report.pdf

Rhee, H.-S., Kim, C., & Ryu, Y. U. (2009). Self-efficacy in Information Security: Its Influence on End Users' Information Security Practice Behavior. *Computers & Security*, *28*(8), 816–826.

Rhee, H.-S., Ryu, Y. U., & Kim, C.-T. (2005). *I Am Fine but You Are Not: Optimistic Bias and Illusion of Control on Information Security*. 381–394. Las Vegas, Nevada.

Robson, C. (2002). *Real World Research: A Resource for Social Scientists and Practitioner-Researchers* (2nd ed.). Oxford: Blackwell.

Rogers, R. W. (1975). A Protection Motivation Theory of Fear Appeals and Attitude Change. *The Journal of Psychology*, *91*(1), 93–114.

Rogers, R. W. (1983). Cognitive and Physiological Processes in Fear Appeals and Attitude Change: A Revised Theory of Protection Motivation. In J. T. Cacioppo & R. E. Petty (Eds.), *Social Psychophysiology*. New York: Guilford Press.

Rusch, J. J. (1999). *The Social Engineering of Internet Fraud*. Presented at the INET'99 Conference, San Jose, CA.

Saunders, M., Lewis, P., & Thornhill, A. (2009). *Research Methods for Business Students* (5th ed.). Prentice Hall.

Scheeres, J. W. (2008). *Establishing the Human Firewall: Reducing an Individual's Vulnerability to Social Engineering Attacks* (Master of Science in Systems Engineering). Air Force Institute of Technology, Wright-Patterson Air Force Base, Ohio.

Schneier, B. (2000). *Secrets and Lies: Digital Security in a Networked World*. Retrieved from http://books.google.co.ke/books?id=z_7CAjmql6kC&num=10

Serianu, C. T. I. T., & USIU, C. for I. R. and I. (2017). *Africa Cyber Security Report 2017: Demystifying Africa's Cyber Security Poverty Line* (p. 88). Retrieved from Serianu Limited website: www.serianu.com/resources.html

Serianu, C. T. I. T., & USIU-A, C. for I. R. and I. (2014). *Kenya Cyber Security Report 2014: Rethinking Cyber Securityâ€ " An Integrated Approach: Processes, Intelligence and Monitoring* (p. 42). Retrieved from Serianu Limited website: www.serianu.com/resources.html

Serianu, C. T. I. T., & USIU-A, C. for I. R. and I. (2015). *Kenya Cyber Security Report 2015: Achieving Enterprise Cyber Resilience Through Situational Awareness* (p. 52). Retrieved from Serianu Limited website: www.serianu.com/resources.html

Serianu, C. T. I. T., & USIU-A, C. for I. R. and I. (2016). *Africa Cyber Security Report 2016: Achieving Cyber Security Resilience* (p. 72). Retrieved from Serianu Limited website: www.serianu.com/resources.html

Serianu, C. T. I. T., & USIU-A, C. for I. R. and I. (2017). *Kenya Cyber Security Report 2017: Demystifying Africa's Cyber Security Poverty Line* (p. 80). Retrieved from Serianu Limited website: www.serianu.com/resources.html

Serianu, C. T. I. T., USIU-A, C. for I. R. and I., & Paladion, P. T. (2016). *Kenya Cyber Security Report 2016: Achieving Cyber Security Resilience* (p. 51). Retrieved from Serianu Limited website: www.serianu.com/resources.html

Sheng, S., Holbrook, M., Kumaraguru, P., Cranor, L. F., & Downs, J. S. (2010). Who Falls for Phish? A Demographic Analysis of Phishing Susceptibility and Effectiveness of Interventions. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 373–382. ACM.

Sheng, S., Magnien, B., Kumaraguru, P., Acquisti, A., Cranor, L. F., Hong, J., & Nunge, E. (2007). Anti-Phishing Phil: The Design and Evaluation of a Game That Teaches People Not to Fall for Phish. *Proceedings of the 3rd Symposium on Usable Privacy and Security*, 88–99. ACM.

Silowash, G., Cappelli, D. M., Moore, A. P., Trzeciak, R., Shimeall, T. J., & Flynn, L. A. (2012, December). *Common Sense Guide to Mitigating Insider Threats*. Retrieved from http://www.sei.cmu.edu/library/abstracts/reports/12tr012.cfm

Siponen, M., & Vance, A. (2010). Neutralization: New insights into the problem of employee information systems secuirty policy violations. *MIS Quarterly*, *34*(3).

Straub, D. W. (1989). Validating Instruments in MIS Research. *MIS Quarterly*, *13*(2), 147–169. https://doi.org/10.2307/248922

Straub, D. W. (1990). Effective IS security: An empirical study. *Information Systems Research*, *1*(3), 255–276.

Straub, D. W., Boudreau, M.-C., & Gefen, D. (2004). Validation Guidelines for IS Positivist Research. *The Communications of the Association for Information Systems*, *13*(1), 63.

Sutton, R. I., & Staw, B. M. (1995). What Theory is Not. *Administrative Science Quarterly*, *40*(3), 371–384.

Tetri, P., & Vuorinen, J. (2013). Dissecting Social Engineering. *Behaviour & Information Technology*, *32*(10), 1014–1023. https://doi.org/10.1080/0144929X.2013.763860

Theoharidou, M., Kokolakis, S., Karyda, M., & Kiountouzis, E. (2005). The Insider Threat to Information Systems and the Effectiveness of ISO 17799. *Computers & Security*, *24*(6), 472–484.

Tsai, H. S., Jiang, M., Alhabash, S., LaRose, R., Rifon, N. J., & Cotten, S. R. (2016). Understanding Online Safety Behaviors: A Protection Motivation Theory Perspective. *Computers & Security*, *59*, 138–150.

Tsow, A., & Jakobsson, M. (2007). *Deceit and Deception: A Large User Study of Phishing*. Indiana University.

Urbano, J., Rocha, A. P., & Oliveira, E. (2013). A Socio-Cognitive Perspective of Trust. In *Agreement Technologies* (pp. 419–429). Springer, Dordrecht.

Vance, A., Siponen, M., & Pahnila, S. (2012). Motivating IS Security Compliance: Insights from Habit and Protection Motivation Theory. *Information & Management*, *49*(3), 190–198.

Verizon. (2015, April). *2015 Data Breach Investigations Report (DBIR)*. Retrieved from http://news.verizonenterprise.com/2015/04/2015-verizon-dbir-report-security/

Verizon. (2016, April). *2016 Data Breach Investigations Report (DBIR)*. Retrieved from http://www.verizonenterprise.com/resources/reports/rp_DBIR_2016_Report_en_xg.pdf

Verizon. (2017). *2017 Data Breach Investigations Report (DBIR)* (p. 76). Retrieved from http://www.verizonenterprise.com/verizon-insights-lab/dbir/2017/

Verizon. (2018). *2018 Data Breach Investigations Report (DBIR)* (p. 76). Retrieved from http://www.verizonenterprise.com/verizon-insights-lab/dbir/2017/

Vishwanath, A., Harrison, B., & Ng, Y. J. (2018). Suspicion, Cognition, Automaticity Model (SCAM) of Phishing Susceptibility. *Communication Research*, *45*(8), 1146–1166. https://doi.org/10.1177/0093650215627483

Vishwanath, A., Herath, T., Chen, R., Wang, J., & Rao, R. H. (2011). Why do people get phished? Testing individual differences in phishing vulnerability within an integrated, information processing model. *Decision Support Systems*, *51*(3), 576–586.

Waleed, A.-G. (2016). Extending Protection Motivation Theory to Understand Security Determinants of Anti-virus Software Usage on Mobiles Devices. *International Journal of Computers*, *10*.

Wang, J., Herath, T., Chen, R., Vishwanath, A., & Rao, R. H. (2012). Phishing Susceptibility: An Investigation Into the Processing of a Targeted Spear Phishing Email. *IEEE Transactions on Professional Communication*, *55*(4), 345–362.

Waqas. (2016, April 28). Anonymous Leaks 1TB of Data from Kenya’s Ministry of Foreign Affairs. Retrieved September 12, 2016, from Hackread website: https://www.hackread.com/anonymous-hacks-kenya-ministry-foreign-affairs/

Weinstein, N. D. (1993). Testing Four Competing Theories of Health-Protective Behavior. *Health Psychology*, *12*(4), 324–333.

Weston, R., & Gore, P. A. (2006). A Brief Guide to Structural Equation Modeling. *The Counseling Psychologist, 34*(5), 719–751.

Wiener, N. (1948). *Cybernetics: Control and Communication in the Animal and the Machine*. Cambridge, MA: MIT Press.

Williams, E. J., Hinds, J., & Joinson, A. N. (2018). Exploring susceptibility to phishing in the workplace. *International Journal of Human-Computer Studies*, *120*, 1–13.

Williams, E. J., & Polage, D. (2019). How persuasive is phishing email? The role of authentic design, influence and current events in email judgements. *Behaviour & Information Technology, 38*(2), 184–197. https://doi.org/10.1080/0144929X.2018.1519599

Wolf, E., Harrington, K. M., Clark, S. L., & Miller, M. W. (2013). Sample Size Requirements for Structural Equation Models: An Evaluation of Power, Bias, and Solution Propriety. *Educational and Psychological Measurement*, *73*, 913–934. https://doi.org/10.1177/0013164413495237

Woon, I., Tan, G.-W., & Low, R. (2005). *A Protection Motivation Theory Approach to Home Wireless Security*. Presented at the International Conference on Information Systems (ICIS).

Workman, M. (2007). Gaining Access with Social Engineering: An Empirical Study of the Threat. *Information Systems Security*, *16*(6), 315–331.

Workman, M. (2008a). A Test of Interventions for Security Threats from Social Engineering. *Information Management & Computer Security*, *16*(5), 463–483.

Workman, M. (2008b). Wisecrackers: A Theory-Grounded Investigation of Phishing and Pretext Social Engineering Threats to Information Security. *Journal of the American Society for Information Science and Technology*, *59*(4), 662–674.

Workman, M., Bommer, W. H., & Straub, D. W. (2008). Security Lapses and the Omission of Information Security Measures: A threat Control Model and Empirical Test. *Computers in Human Behavior, 24*(6), 2799–2816.

Wu, M., Miller, R. C., & Garfinkel, S. L. (2006). Do Security Toolbars Actually Prevent Phishing Attacks? *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 601–610. Montréal, Québec, Canada: ACM.

Yasin, A., Fatima, R., Liu, L., Yasin, A., & Wang, J. (2019). Contemplating Social Engineering Studies and Attack Scenarios: A Review Study. *Security and Privacy*, *e73*. https://doi.org/10.1002/spy2.73

Zhang, Y., Egelman, S., Cranor, L., & Hong, J. (2007). Phinding Phish: Evaluating Anti-Phishing Tools. *Proceedings of the 14th Annual Network and Distributed System Security Symposium (NDSS 2007)*. Presented at the San Diego, CA. San Diego, CA.

Zimmerman, R. A., Friedman, G. M., Munshi, D. C., Richmond, D. A., & Jaros, S. L. (2018). *Modeling Insider Threat from the Inside and Outside: Individual and Environmental Factors Examined Using Event History Analysis* (No. PERSEREC-TR-18-14, OPA-2018-065). Retrieved from Defense Personnel and Security Research Center DoD Agencies United States website: https://apps.dtic.mil/docs/citations/AD1058522

APPENDIX A: DATA COLLECTION LETTER

UNIVERSITY OF NAIROBI
COLLEGE OF BIOLOGICAL AND PHYSICAL SCIENCES
SCHOOL OF COMPUTING AND INFORMATICS

Telephone:	4447870/ 4444919/44446544	P. O. Box 30197
Telegrams:	"Varsity" Nairobi	Nairobi
Telefax:	254-020-4447870	Kenya
Email:	director-sci@uonbi.ac.ke	

1 August 2014

TO WHOM IT MAY CONCERN

Dear Sir/Madam,

RE: SUBJECT: DOCTORAL RESEARCH BY PAULA M. W. MUSUVA

Paula Mwikali Wasua Musuva of ID number 22584736 is a Doctoral student at the University of Nairobi School of Computing and Informatics student number P80/92723/2013. Her research proposal and data collection methodology and instruments have been approved and she is now embarking on her data collection phase.

The purpose of this study is to *"Investigate the Insider Threat to Information Systems Security in Kenyan Commercial Banks".* The research has been designed to collect information on information security practices of Kenyan Commercial Banks and their employees.

Participant responses will be confidential and data from this research will be reported in collective terms and in ways that will not be personally identifiable to these organization or their employees.

If you have any questions about this research, feel free to contact me or the Principal Investigator at pmusuva@gmail.com or (+254) 020 3606152 or (+254)722 540332

Your contribution and support is greatly appreciated.

Yours sincerely,

School of Computing & Informatics
University of NAIROBI
P. O. Box 30197
NAIROBI

CHRISTOPHER CHEPKEN, PHD.
DOCTORAL SUPERVISOR
SCHOOL OF COMPUTING AND INFORMATICS
UNIVERSITY OF NAIROBI

CCK/jsn

USIU AFRICA

United States International University-Africa

27th August, 2015

Paula Musuva Kigen
Lecturer, School of Science and Technology
United States International University (USIU)
P.O. Box 14634, Code 00800, Nairobi, Kenya

Dear Ms. Musuva,

Ref: Research Approval Letter

Following your request to carry out research at USIU dated August 26, 2015 on the topic ***"Modeling the Unintentional Insider Threat to Information Systems Security,"*** the university's research office has authorized you to pursue your research.

We hope that you would be able to conduct your research successively and we wish you the very best in your efforts towards completion of the project. USIU-A places great importance on dissemination of data and research output and would therefore require a copy of the completed research report.

Sincerely,

Francis W. Wambalaba, Ph.D., AICP
Associate Deputy Vice Chancellor Academics:
Research
United States International University
P.O. Box 14634, Nairobi, Kenya, 00800
fwambalaba@usiu.ac.ke
PH. + 254 20 3606442

APPENDIX B: INSTITUTIONAL REVIEW BOARD APPROVAL

USIU AFRICA

United States
International
University-Africa

USIU-A Institutional Review Board (IRB)

23rd June 2016

USIU-A/IRB/16/F06

Paula Musuva Kigen,
Lecturer, School of Science and Technology,
United States International University (USIU),
P.O. Box 14634,
Nairobi, Kenya.

<u>IRB-RESEARCH APPROVAL.</u>

The USIU-A IRB has reviewed and granted ethical approval for the research proposal titled *"Modeling the Unintentional Insider Threat to Information Systems Security".* The approval is for **six months** from the date of IRB. Please submit a completed copy of the study to the IRB office, soft copy is acceptable.
You are advised to follow the approved methodology and report to the IRB any serious, unexpected and related adverse events and potential unanticipated problems involving risks to subjects or others.

Should you or study participants have any queries regarding IRB's consideration of this project, please contact irb@usiu.ac.ke.

Dr. Carol J. Watson,
Chair | IRB | USIU-Africa
cwatson@usiu.ac.ke
Office 20 3606 303
Cell +254 70101 7099

CC: Research Office

APPENDIX C: INDEX HTML PAGE SOURCE CODE

```
<?php
        session_start();

        // Initialize variables
        $username="";
        $email="";
        $passwordErr="";
        $nameErr ="";
        $emailErr ="";
        $passwordErr ="";
        $isValidUsername=0;
        $isValidEmail = 0;
        $isValidPassword = 0;

        function test_input($data) {
                $data = trim($data);
                $data = stripslashes($data);
                $data = htmlspecialchars($data);
                return $data;
        }

        if($_SERVER["REQUEST_METHOD"] == "POST") {
                // Form submitted

                //-------------------------Form Validation Start---------------------//
                if (empty($_POST["username"])) {
                        $nameErr = "Name is required";
                        $isValidUsername = 0;
                } else {
                        $username = test_input($_POST["username"]);
                        if (!preg_match("/^[a-zA-Z ]*$/",$username)) {
                                $nameErr = "Only letters and white space allowed";
                                $isValidUsername = 0;
                        }
                        else {
                                $isValidUsername = 1;
                        }
                }

                if (empty($_POST["email"])) {
                        $emailErr = "E-mail is required e.g. username@uni.ac.ke";
                        $isValidEmail = 0;
                } else {
                        $email = test_input($_POST["email"]);
                        $regex = '/^[_a-z0-9-]+(\.[_a-z0-9-]+)*@[a-z0-9-]+(\.[a-z0-9-]+)*(\.[a-
        z]{2,4})$/';
                        if (!preg_match($regex, $email)) {
                                $emailErr = "$email is not a valid email address";
                                $isValidEmail = 0;
                        }
                        else {
                                $isValidEmail = 1;
                        }
                }

                if (empty($_POST["password"])) {
                        $passwordErr = "Password is required";
                        $isValidPassword = 0;
                } else {
                        $password = md5($_POST["password"]);
                        $isValidPassword = 1;
                }

                //-------------------------Form Validation End---------------------//

                //-------------------------Database Connection Start---------------------//
                if ($isValidUsername && $isValidEmail && $isValidPassword){
                        //Set up connection to database
                        define('DB_SERVER', 'SERVER_NAME');
                        define('DB_USERNAME', 'USER_NAME');
                        define('DB_PASSWORD', 'PASSWORD');
                        define('DB_DATABASE', 'DB_NAME');
                        $db = mysqli_connect(DB_SERVER,DB_USERNAME,DB_PASSWORD,DB_DATABASE);

                        if (!$db) {
                                die("Connection failed: " . mysqli_connect_error());
                        }

                        //mysqli_real_escape_string used to prevent SQLi
                        $username = mysqli_real_escape_string($db,$username);
                        $email = mysqli_real_escape_string($db,$email);

                        //No password stored to protect users
                        $sql = "INSERT INTO responses (`names`, `email`) VALUES
                        ('$username','$email')";

                        if (mysqli_query($db,$sql)){
                                echo "Your email quota has been increased to 4GB";
                                echo '<script
                                type="text/javascript">window.location.href="http://www.UNI.ac.k
                                e";</script>';
                                }
                                else {
                                echo "Error: " . $sql . "<br>" . mysqli_error($db);
                                }

                        mysqli_close($db);
                }
                //-------------------------Database Connection End---------------------//

        }

?>

//-------------------------HTML5 Index Page Start---------------------//

<!doctype html>
<html lang="en">

<head>
        <meta charset="utf-8">
        <title>E-mail Quota</title>
        <link rel="stylesheet" type="text/css" href="stylesheet.css">
</head>

<body topmargin='0' bottommargin='0' leftmargin='0' rightmargin='0' marginwidth='0'
        marginheight='0' Onload="fillEmail()">

<br>

<center>

<table border=0 cellpadding=5 cellspacing=5 width='900' height='300'>

<tr
<td align=center bgcolor=white>
<table border=0 cellpadding=5 cellspacing=5 bgcolor=#ffffff width='100%'>
```

```
<tr valign=top>
<td colspan=3><h1 align=center >E-mail Quota Extension</h1></td>
</tr>

<tr valign=top>
<td align=center><img src="images/logo.jpg" border=0></td>
<td>

<table cellpadding=0 cellspacing=0 border=0>

<tr>
<td>

<table border=0 cellpadding=2 cellspacing=5 width='100%'>

<form method=post action="<?php echo htmlspecialchars($_SERVER["PHP_SELF"]);?>">

<tr>
<td>Full Names: </td>
<td><input type=text name=username class=nicefield size=40 maxlength=255 value=<?php echo
        $username;?>>
<br><span class="error"><?php echo $nameErr;?></span>
</td>
</tr>

<tr>
<td>E-mail address: </td>
<td><input type=text name=email class=nicefield size=40 maxlength=255 value=<?php echo $email;?>>
<br><span class="error"><?php echo $emailErr;?></span>
</td>
</tr>

<tr>
<td>Password: </td>
<td><input type=password name=password class=nicefield size=40 maxlength=255>
<br><span class="error"><?php echo $passwordErr;?></span>
</td>
</tr>

<tr>
<td>Increase Quota (4GB): </td>
<td><input type=checkbox checked name=checkboxQuota class=nicecheckbox></td>
</tr>

<tr>
<td></td>
<td align=left><input type=submit name=btnsubmit value='Submit' class=nicebutton></td>
</form></tr>
</table>
</td>
</tr>

</table>
</center>
</td>
</tr>
</table>
</center>
</body>
</html>
//-------------------------HTML5 Index Page End---------------------//
```

APPENDIX D: CASCADING STYLE SHEETS CODE

```
tr, td, p {
        font-family: Segoe, Tahoma, Arial, Helvetica, Sans-serif;
        font-size: 14px;
        color: #000000;
        letter-spacing: 0px;
        height: 35px;
        margin-top: 5px;
        margin-left: 0px;
        margin-right: 0px;
        margin-bottom: 5px;
        margin: 0px;
}

h1 {
        font-family: Segoe, Tahoma, Arial, Helvetica, Sans-serif;
        font-size: 18px;
        font-weight: bold;
        letter-spacing: -1px;
        color: navy;
        padding: 0;
        margin: 0px 0 0 0;
        line-height: 1em;
        padding-top: 3px;
}
.error {
        font-size: 11px;
        color: red;
}
.nicebutton {
        font-size: 14px;
        height: 35px;
        width: 140px;
        color: black;
}

.nicefield {
        font-size: 14px;
        color: #000000;
        height: 30px;
}

.nicecheckbox {
        height: 20px;
        width: 20px;
        color: #000000;
}
```

APPENDIX E: BACKGROUND SCRIPT SOURCE CODE

```php
<?php
session_start();

//-------------------------User Detection Start---------------------//
$user_agent    =   $_SERVER['HTTP_USER_AGENT'];

function getOS() {
        global $user_agent;
        $os_platform   =   "Unknown OS Platform";
        $os_array      =   array(
                                '/windows nt 10/i'    =>  'Windows 10',
                                '/windows nt 6.3/i'    =>  'Windows 8.1',
                                '/windows nt 6.2/i'    =>  'Windows 8',
                                '/windows nt 6.1/i'    =>  'Windows 7',
                                '/windows nt 6.0/i'    =>  'Windows Vista',
                                '/windows nt 5.2/i'    =>  'Windows Server 2003/XP x64',
                                '/windows nt 5.1/i'    =>  'Windows XP',
                                '/windows xp/i'        =>  'Windows XP',
                                '/windows nt 5.0/i'    =>  'Windows 2000',
                                '/windows me/i'        =>  'Windows ME',
                                '/win98/i'             =>  'Windows 98',
                                '/win95/i'             =>  'Windows 95',
                                '/win16/i'             =>  'Windows 3.11',
                                '/macintosh|mac os x/i' =>  'Mac OS X',
                                '/mac_powerpc/i'       =>  'Mac OS 9',
                                '/linux/i'             =>  'Linux',
                                '/ubuntu/i'            =>  'Ubuntu',
                                '/iphone/i'            =>  'iPhone',
                                '/ipod/i'              =>  'iPod',
                                '/ipad/i'              =>  'iPad',
                                '/android/i'           =>  'Android',
                                '/blackberry/i'        =>  'BlackBerry',
                                '/webos/i'             =>  'Mobile'
                                        );
                foreach ($os_array as $regex => $value) {
                        if (preg_match($regex, $user_agent)) {
                                $os_platform   =   $value;
                        }
                }
                return $os_platform;
        }

function getBrowser() {
        global $user_agent;
        $browser       =   "Unknown Browser";
        $browser_array =   array(
                                '/msie/i'      =>  'Internet Explorer',
                                '/firefox/i'   =>  'Firefox',
                                '/safari/i'    =>  'Safari',
                                '/chrome/i'    =>  'Chrome',
                                '/edge/i'      =>  'Edge',
                                '/opera/i'     =>  'Opera',
                                '/netscape/i'  =>  'Netscape',
                                '/maxthon/i'   =>  'Maxthon',
                                '/konqueror/i'  =>  'Konqueror',
                                '/mobile/i'    =>  'Handheld Browser'
                                        );
                foreach ($browser_array as $regex => $value) {
                        if (preg_match($regex, $user_agent)) {
                                $browser    =   $value;
                        }
                }
                return $browser;
        }
        function getRealUserIp(){
                switch(true){
                  case (!empty($_SERVER['HTTP_X_REAL_IP'])) : return $_SERVER['HTTP_X_REAL_IP'];
                  case (!empty($_SERVER['HTTP_CLIENT_IP'])) : return $_SERVER['HTTP_CLIENT_IP'];
                  case (!empty($_SERVER['HTTP_X_FORWARDED_FOR'])) : return
        $_SERVER['HTTP_X_FORWARDED_FOR'];
                  default : return $_SERVER['REMOTE_ADDR'];
                }
        }

        $user_ip = getRealUserIp();
        $user_browser = getBrowser();
        $user_os = getOS();
        $hostname = gethostbyaddr($_SERVER['REMOTE_ADDR']);


        //-------------------------DB Connection---------------------//
        //Only executes if email variable is provided from email link
        if (isset($_GET['email'])) {
                define('DB_SERVER', 'SERVER_NAME');
                define('DB_USERNAME', 'USER_NAME');
                define('DB_PASSWORD', 'PASSWORD');
                define('DB_DATABASE', 'DB_NAME');

                $db = mysqli_connect(DB_SERVER,DB_USERNAME,DB_PASSWORD,DB_DATABASE);

                if (!$db) {
                        die("Connection failed: " . mysqli_connect_error());
                }

                $email = mysqli_real_escape_string($db,$_GET['email']);

                //set email session variable to us in form
                $_SESSION['email'] = $email;

                //SQL Query
                $sql = "INSERT INTO TABLE_NAME (`email`, `IP`, `Browser`, `OS`, `Hostname`,
                `UserAgent`) VALUES
                ('$email','$user_ip','$user_browser','$user_os','$hostname','$user_agent')";

                if (mysqli_query($db,$sql)){
                        echo "Opening...<br>";
                }
                else {
                        echo "Error: " . $sql . "<br>" . mysqli_error($db);
                }
        }

        echo '<script
        type="text/javascript">window.location.href="http://usiu.or.ke/email/";</script>';

        mysqli_close($db);
?>
```

APPENDIX F: DATA COLLECTION QUESTIONNAIRE

INFORMATION SECURITY SURVEY

Dear respondent,

I am Paula Musuva-Kigen, a faculty member at [redacted] in the School of Science and Technology and also a student at the University of Nairobi School of Computing and Informatics pursuing a Doctor of Philosophy in Information Systems. As part of my degree requirements I am conducting a research study. The purpose of this study is to determine the factors that contribute to the unintentional insider threat to information systems security.

Kindly fill in the questionnaire as accurately and honestly as possible. It should take approximately 15 minutes to do so.
For each question, *please tick one* option that best fits your answer. You may also write responses in the spaces provided.

Your responses will be confidential and data from this research will be reported in collective terms and in ways that will not be personally identifiable to you.

If you have any questions about this research, feel free to contact me at pmusuva@[redacted] or (020) [redacted].

Please sign below to indicate that you are above 18 years of age, you have been informed of the study and that you give your written consent to participate in the study.

Signature: _____________________________ Date: ____________________ Participant Number: _________________

SECTION A: BACKGROUND INFORMATION

1. What is your gender?
   □ Male □ Female

2. What is your age in years?
   □ less than 18 years □ 18 - 25 years □ 26 - 35 years
   □ 36 - 45 years □ 46 - 55 years □ above 55 years

3. What is the highest level of education you have *completed*?
   □ Primary School
   □ High School
   □ Diploma
   □ Undergraduate Degree (Bachelor's)
   □ Graduate Degree (Master's)
   □ Doctoral Degree (PhD)

4. What is your role at the university?
   □ Student □ Staff
   □ Faculty/Lecturer □ Other: ___________________

5. In which school or department are you in?
   □ Chandaria School of Business
   □ School of Humanities and Social Sciences
   □ School of Pharmacy and Health Sciences
   □ School of Science and Technology
   □ Quality Assurance
   □ Finance and Administration
   □ Student Affairs
   □ Library
   □ Cafeteria
   □ Maintenance
   □ Legal Affairs
   □ Other: ___________________

6. How many years have you been at the university? _________ years

7. Which year did you first use the internet?
   □ before 1991 □ 1991-1995 □ 1996 -2000 □ 2001-2005 □ 2006-2010 □ after 2010

8. How many hours do you spend on the internet in a day?
   □ less than 5 □ 5-10 □ 11-15 □ 16-20 □ 21-24

9. How would you rate your computer skills?

□ Low	□ Basic	□ Intermediate	□ Advanced	□ Expert
(little or no skill; requiring a lot of assistance to perform tasks on a computer)	(I can navigate a computer and perform simple tasks such as prepare documents, print and respond to emails)	(in addition to basic tasks I can prepare and analyze data on spreadsheets, make presentations with little or no assistance)	(in addition to intermediate tasks, I can change configuration settings, customize applications, backup and manage personal data)	(in addition to advanced tasks, I can write applications, audit and secure computer systems, troubleshoot and solve computer problems

RISK PROPENSITY

To what extent do you agree with the following statements about your risk propensity?

RP1: I like taking risks

strongly disagree ○—○—○—○—○ strongly agree
1 2 3 4 5

RP2: People say I am a risk taker

strongly disagree ○—○—○—○—○ strongly agree
1 2 3 4 5

RP3: I sometimes take risks that could threaten my safety

strongly disagree ○—○—○—○—○ strongly agree
1 2 3 4 5

ONLINE SERVICES

To what extent do you use the following online services?

OS1: Email not at all ○—○—○—○—○ very great extent
1 2 3 4 5

OS2: Social Media not at all ○—○—○—○—○ very great extent
1 2 3 4 5

OS3: Online Shopping not at all ○—○—○—○—○ very great extent
1 2 3 4 5

OS4: Online Banking not at all ○—○—○—○—○ very great extent
1 2 3 4 5

PRIOR VICTIMIZATION

Have you ever experienced the following online threats in the past?

PV1: **Scam** (e.g. receiving an email that convinces you to reveal personal details or send money to a falsified recipient)	□ Yes	□ No	□ I Don't Know
PV2: **Online Account Hijacking** (e.g. someone takes over your online account and sends messages pretending to be you)	□ Yes	□ No	□ I Don't Know
PV3: **Identity Theft** (e.g. someone opens an account pretending to be you and takes up your persona online)	□ Yes	□ No	□ I Don't Know
PV4: **Credit/Debit Card Fraud** (e.g. someone uses your credit/debit card to make payments without your knowledge or consent)	□ Yes	□ No	□ I Don't Know
PV5: **Malicious software infection** (e.g. your computer is infected by viruses thus degrading its performance and compromising the confidentiality and integrity of your data)	□ Yes	□ No	□ I Don't Know

SECTION B: INFORMATION SECURITY KNOWLEDGE

Please indicate what the following words mean with regards to information security. Do so to the best of your knowledge without consulting or referring to any material. For each of the words, fill in the blank with one letter indicating your choice from the list of definitions. e.g. Malware ___I___ **(You can select a letter more than once)**

KQ1: Phishing _______

KQ2: Social Engineering _______

KQ3: URL _______

KQ4: Certificate _______

KQ5: Spoofing _______

KQ6: Domain _______

A. I have never seen this word before
B. I have seen this word before but I don't know what it means
C. A file used to identify websites and encrypt data
D. Manipulating people to compromise the security of their systems
E. A name that identifies an organization's resources on the internet
F. Forging the identity of a trusted entity
G. Impersonation commonly through email that tricks people into sharing sensitive information
H. A term for insecure websites
I. Malicious Software
J. A web address

PERCEIVED SEVERITY

Please rate how bad you think the consequences of the following actions could be on the internet:

PS1: Opening a suspicious email	not at all	○ 1	○ 2	○ 3	○ 4	○ 5	very great extent
PS2: Opening a suspicious attachment	not at all	○ 1	○ 2	○ 3	○ 4	○ 5	very great extent
PS3: Clicking a suspicious hyperlink	not at all	○ 1	○ 2	○ 3	○ 4	○ 5	very great extent
PS4: Loading a suspicious website	not at all	○ 1	○ 2	○ 3	○ 4	○ 5	very great extent
PS5: Filling out personal details on a website	not at all	○ 1	○ 2	○ 3	○ 4	○ 5	very great extent
PS6: Sharing my [redacted] username and password	not at all	○ 1	○ 2	○ 3	○ 4	○ 5	very great extent

EMAIL LOAD

EL: How many emails do you receive in your official [redacted] email account in a day?

□ less than 10 □ 11-20 □ 21-30 □ 31-40 □ 41-50 □ more than 50

EMAIL RESPONSE

To what extent do you agree with the following statements?

ER1: I *read* all emails I receive in my official [redacted] email account

strongly disagree ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 strongly agree

ER2: I *respond* to all emails I need to in my official [redacted] email account

strongly disagree ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 strongly agree

RESPONSIBLE

RES1: I am answerable to communications I receive on my [redacted] email account

strongly disagree ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 strongly agree

RES2: I am in control of the day-to-day operation of my [redacted] email account

strongly disagree ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 strongly agree

RES3: I consider myself responsible for my [redacted] email account

strongly disagree ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 strongly agree

ABILITY TO PROCESS

AP1: I am usually clear-headed when reading and responding to emails

strongly disagree ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 strongly agree

AP2: I am usually able to concentrate when reading and responding to emails

strongly disagree ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 strongly agree

AP3: I am usually relaxed when reading and responding to emails

strongly disagree ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 strongly agree

DISTRACTION

DIST1: There is usually a lot of activity going on around me when reading and responding to emails

strongly disagree ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 strongly agree

DIST2: I usually multi-task when reading and responding to emails

strongly disagree ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 strongly agree

DIST3: I tend to be distracted when reading and responding to emails

strongly disagree ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 strongly agree

PRESSURE

PRES1: I am usually under pressure to move on to other tasks when reading and responding to emails

strongly disagree ○ ○ ○ ○ ○ strongly agree
1 2 3 4 5

PRES2: I usually have a sense of urgency when reading and responding to emails

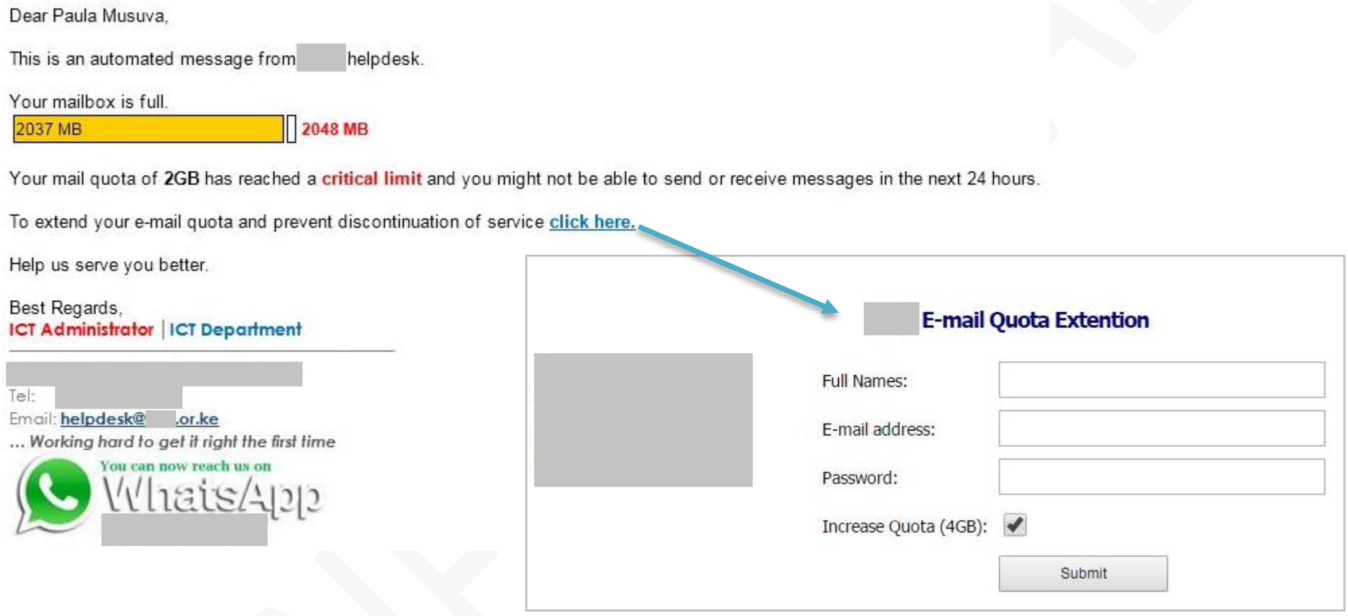
strongly disagree ○ ○ ○ ○ ○ strongly agree
1 2 3 4 5

PRES3: I tend to rush through my emails

strongly disagree ○ ○ ○ ○ ○ strongly agree
1 2 3 4 5

SECTION C: PHISHING SCENARIO

You have been asked to respond to this questionnaire because you interacted with an email sent to you from the domain [redacted].or.ke that impersonated [redacted].ac.ke with the subject **"Your mailbox is full"** as shown below. It directed you to a form which asked you to fill in your personal details (names, email address and password). This scenario was designed to resemble a common threat that many users are exposed to.



Dear Paula Musuva,

This is an automated message from [redacted] helpdesk.

Your mailbox is full.

2037 MB 2048 MB

Your mail quota of **2GB** has reached a **critical limit** and you might not be able to send or receive messages in the next 24 hours.

To extend your e-mail quota and prevent discontinuation of service click here.

Help us serve you better.

Best Regards,
ICT Administrator | ICT Department

[redacted]
Tel: [redacted]
Email: helpdesk@[redacted].or.ke
... Working hard to get it right the first time
You can now reach us on WhatsApp
[redacted]

[redacted]

[redacted] E-mail Quota Extention

Full Names:

E-mail address:

Password:

Increase Quota (4GB): ☑

Submit

Please answer the following questions as relates to your interaction with the email

OB1: Did you read this email? □ Yes □ No

OB2: Did you click the link labelled "click here" on this email? □ Yes □ No

OB3: Did you fill in the form presented on the website? □ Yes □ No

NOTE: If you did not read this email or access the website, please answer the questionnaire based on your evaluation of the described scenario. You could also respond based on your experience with similar emails.

MOTIVATED TO PROCESS

MP1: After reading the subject heading "Your mailbox is full" I had to open the email

strongly disagree ○ ○ ○ ○ ○ strongly agree
1 2 3 4 5

MP2: I did not want to ignore a message regarding my mailbox being full

strongly disagree ○ ○ ○ ○ ○ strongly agree
1 2 3 4 5

MP3: I was committed to attending to the issue presented regarding my mailbox being full

strongly disagree ○ ○ ○ ○ ○ strongly agree
1 2 3 4 5

INVOLVEMENT

INV1: The email seemed very relevant to me

strongly disagree ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 strongly agree

INV2: The email seemed very important to my work/studies

strongly disagree ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 strongly agree

INV3: The email seemed very applicable to my current situation

strongly disagree ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 strongly agree

ELABORATION

ELAB1: I made conscious effort to evaluate the email/website

strongly disagree ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 strongly agree

ELAB2: I took time to evaluate the email/website

strongly disagree ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 strongly agree

ELAB3: I carefully evaluated the email/website

strongly disagree ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 strongly agree

QUALITY OF ARGUMENT

QA1: I carefully scrutinized the email message before responding

strongly disagree ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 strongly agree

QA2: I reasoned through the explanation given in the email before responding

strongly disagree ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 strongly agree

QA3: I examined the reasons given in the email before responding

strongly disagree ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 strongly agree

EMOTIONS

EM1: Reading the email invoked an emotion in me (e.g. fear, anxiety)

strongly disagree ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 strongly agree

EM2: I responded to this email so that I would not get into trouble

strongly disagree ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 strongly agree

EM3: I would have felt guilty for not responding to the email

strongly disagree ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 strongly agree

PERSUASIVE CUES

Please rate to which extent the following components of the email/website influenced your response:

PC1: Source credibility (i.e. ICT administrator)	not at all	○ 1	○ 2	○ 3	○ 4	○ 5	very great extent
PC2: Personalized Greeting	not at all	○ 1	○ 2	○ 3	○ 4	○ 5	very great extent
PC3: Offer to extend your mail quota	not at all	○ 1	○ 2	○ 3	○ 4	○ 5	very great extent
PC4: Warning that your email service would be discontinued	not at all	○ 1	○ 2	○ 3	○ 4	○ 5	very great extent
PC5: Urgency to respond within 24 hours	not at all	○ 1	○ 2	○ 3	○ 4	○ 5	very great extent
PC6: Resemblance to other [redacted] emails	not at all	○ 1	○ 2	○ 3	○ 4	○ 5	very great extent
PC7: Resemblance to other [redacted] websites	not at all	○ 1	○ 2	○ 3	○ 4	○ 5	very great extent

DETECTION CUES

DC1: I know how to reveal hyperlinks hidden behind text to detect such threats

strongly disagree ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 strongly agree

DC2: I know how to analyze web addresses to detect such threats

strongly disagree ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 strongly agree

DC3: I know how to analyze web certificates to detect such threats

strongly disagree ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 strongly agree

Generally, to what extent did you think that:

BI1: the email was trustworthy	not at all	○ 1	○ 2	○ 3	○ 4	○ 5	very great extent
BI2: the website was trustworthy	not at all	○ 1	○ 2	○ 3	○ 4	○ 5	very great extent

TRUST DETERMINANTS

To what extent did you use the following characteristics or techniques to determine the trustworthiness of the email/website?

DT1: Consistency in logo, colors, look and feel	not at all	○ 1	○ 2	○ 3	○ 4	○ 5	very great extent
DT2: Grammar and Spelling	not at all	○ 1	○ 2	○ 3	○ 4	○ 5	very great extent
DT3: Personalized greeting with your names	not at all	○ 1	○ 2	○ 3	○ 4	○ 5	very great extent
DT4: Content (e.g. reasonableness of the explanation in email and website content)	not at all	○ 1	○ 2	○ 3	○ 4	○ 5	very great extent
DT5: Context (e.g. it was expected in the prevailing circumstances)	not at all	○ 1	○ 2	○ 3	○ 4	○ 5	very great extent
DT6: Email address of the sender	not at all	○ 1	○ 2	○ 3	○ 4	○ 5	very great extent
DT7: Contacting the [redacted] ICT helpdesk	not at all	○ 1	○ 2	○ 3	○ 4	○ 5	very great extent
DT8: Asking someone (e.g. colleague, friend)	not at all	○ 1	○ 2	○ 3	○ 4	○ 5	very great extent
DT9: Web address and hyperlink evaluation	not at all	○ 1	○ 2	○ 3	○ 4	○ 5	very great extent
DT10: Website encryption or padlock icon	not at all	○ 1	○ 2	○ 3	○ 4	○ 5	very great extent
DT11: Website certificate	not at all	○ 1	○ 2	○ 3	○ 4	○ 5	very great extent
DT12: Domain registration information (e.g. from whois)	not at all	○ 1	○ 2	○ 3	○ 4	○ 5	very great extent
DT13: Security tool information (e.g. anti-phishing tool integrated in email/browser)	not at all	○ 1	○ 2	○ 3	○ 4	○ 5	very great extent

THREAT DETECTION

TD1: I could tell this was an online attack

strongly disagree ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 strongly agree

TD2: I could tell someone was trying to deceive me

strongly disagree ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 strongly agree

TD3: I could tell that someone was trying to capture my personal details and password

strongly disagree ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 strongly agree

KNOWLEDGE

KW1: I have sufficient knowledge regarding this type of threat

strongly disagree ○—○—○—○—○ strongly agree
1 2 3 4 5

KW2: I have sufficient knowledge regarding the consequences of this type of threat

strongly disagree ○—○—○—○—○ strongly agree
1 2 3 4 5

KW3: I have sufficient knowledge on how to detect this type of threat

strongly disagree ○—○—○—○—○ strongly agree
1 2 3 4 5

KW4: I have sufficient knowledge on how to respond to this type of threat

strongly disagree ○—○—○—○—○ strongly agree
1 2 3 4 5

THREAT APPRAISAL

TAP1: This threat is a real problem facing internet users

strongly disagree ○—○—○—○—○ strongly agree
1 2 3 4 5

TAP2: This threat can have adverse effects if successfully executed

strongly disagree ○—○—○—○—○ strongly agree
1 2 3 4 5

TAP3: This threat can be devastating if successfully executed

strongly disagree ○—○—○—○—○ strongly agree
1 2 3 4 5

THREAT AVOIDANCE

To what extent do you agree with the following statements regarding your intentions before you took action?

TAV1: My intention was to protect my computer resources

not at all ○—○—○—○—○ very great extent
1 2 3 4 5

TAV1: My intention was to protect my data

not at all ○—○—○—○—○ very great extent
1 2 3 4 5

ATTACK FACTORS

AF1: This scenario presented a well-designed attack

strongly disagree ○—○—○—○—○ strongly agree
1 2 3 4 5

AF2: This scenario presented a convincing attack

strongly disagree ○—○—○—○—○ strongly agree
1 2 3 4 5

AF3: This scenario presented a deceptive attack

strongly disagree ○—○—○—○—○ strongly agree
1 2 3 4 5

COPING APPRAISAL

CA1: I know how to protect myself from such threats

strongly disagree ○—○—○—○—○ strongly agree
1 2 3 4 5

CA2: I am equipped to protect myself from such threats

strongly disagree ○—○—○—○—○ strongly agree
1 2 3 4 5

CA3: I can protect myself from such threats

strongly disagree ○—○—○—○—○ strongly agree
1 2 3 4 5

SELF-EFFICACY

SE1: I could learn to protect myself from similar threats without much assistance

strongly disagree ○—○—○—○—○ strongly agree
1 2 3 4 5

SE2: It would be easy for me to learn security measures to protect myself from similar threats

strongly disagree ○—○—○—○—○ strongly agree
1 2 3 4 5

SE3: I can learn new computer security skills without much difficulty

strongly disagree ○—○—○—○—○ strongly agree
1 2 3 4 5

PERCEIVED VULNERABILITY

PVUL1: The chances of receiving fraudulent emails are high

strongly disagree ○—○—○—○—○ strongly agree
1 2 3 4 5

PVUL2: I am a likely target for online attacks

strongly disagree ○—○—○—○—○ strongly agree
1 2 3 4 5

PVUL3: I am likely to encounter various online attacks

strongly disagree ○—○—○—○—○ strongly agree
1 2 3 4 5

RESPONSE EFFICACY

RE1: Enabling security measures would protect users from similar threats

strongly disagree ○—○—○—○—○ strongly agree
1 2 3 4 5

RE2: Enabling security measures would prevent users from being deceived by similar threats

strongly disagree ○—○—○—○—○ strongly agree
1 2 3 4 5

RE3: Enabling security measures would prevent attackers from successfully launching similar threats

strongly disagree ○—○—○—○—○ strongly agree
1 2 3 4 5

PERCEIVED BENEFIT

PB1: Taking precautions to prevent similar attacks would be worthwhile

strongly disagree ○—○—○—○—○ strongly agree
1 2 3 4 5

PB2: [redacted] would benefit greatly from protecting its systems from similar attacks

strongly disagree ○—○—○—○—○ strongly agree
1 2 3 4 5

PB3: Protecting myself from similar attacks would be beneficial

strongly disagree ○—○—○—○—○ strongly agree
1 2 3 4 5

RESPONSE COST

RC1: Taking precautions to prevent such threats would be an inconvenience

strongly disagree ○—○—○—○—○ strongly agree
1 2 3 4 5

RC2: Taking precautions to prevent such threats would be time consuming

strongly disagree ○—○—○—○—○ strongly agree
1 2 3 4 5

RC3: Taking precautions to prevent such threats would hinder my productivity

strongly disagree ○—○—○—○—○ strongly agree
1 2 3 4 5

ORGANIZATIONAL FACTORS

OF1: [redacted] has equipped me to protect myself from such threats

strongly disagree ○—○—○—○—○ strongly agree
1 2 3 4 5

OF2: [redacted] has put in place security measures to protect me from such threats

strongly disagree ○—○—○—○—○ strongly agree
1 2 3 4 5

OF3: [redacted] has shown me how to protect myself from such threats

strongly disagree ○—○—○—○—○ strongly agree
1 2 3 4 5

POLICIES

POL1: I am required to know a lot about [redacted] information security policies

strongly disagree ○—○—○—○—○ strongly agree
1 2 3 4 5

POL2: I know the regulations outlined in [redacted] information security policies

strongly disagree ○—○—○—○—○ strongly agree
1 2 3 4 5

POL3: [redacted] information security policies can guide me in handling such threats

strongly disagree ○—○—○—○—○ strongly agree
1 2 3 4 5

SECURITY EDUCATION, TRAINING & AWARENESS

SETA1: [redacted] has made me aware of such threats

strongly disagree ○—○—○—○—○ strongly agree
1 2 3 4 5

SETA2: [redacted] has provided me with training on how to handle such threats

strongly disagree ○—○—○—○—○ strongly agree
1 2 3 4 5

SETA3: [redacted] has given me sufficient information regarding such threats

strongly disagree ○—○—○—○—○ strongly agree
1 2 3 4 5

TECHNOLOGY CONTROLS

TC1: [redacted] has equipped me with technology controls that can detect such threats

strongly disagree ○—○—○—○—○ strongly agree
1 2 3 4 5

TC2: [redacted] has equipped me with technology controls that can prevent such threats

strongly disagree ○—○—○—○—○ strongly agree
1 2 3 4 5

TC3: [redacted] has equipped me with technology controls that can protect me from such threats

strongly disagree ○—○—○—○—○ strongly agree
1 2 3 4 5

SECURITY MEASURES

To what extent do you agree that the following security measures would protect someone from such threats?

SM1: Information Security Policy (e.g. explicit guidelines in a university-wide ICT policy)

strongly disagree ○—○—○—○—○ strongly agree
1 2 3 4 5

SM2: Information Security Education, Training or Awareness (e.g. attending anti-phishing training)

strongly disagree ○—○—○—○—○ strongly agree
1 2 3 4 5

SM3: Installing Technology Controls (e.g. email filters, web monitoring, anti-phishing tools)

strongly disagree ○—○—○—○—○ strongly agree
1 2 3 4 5

SM4: Secure operational procedures (e.g. operational guidelines regarding verifying legitimacy of online messages)

strongly disagree ○—○—○—○—○ strongly agree
1 2 3 4 5

SM5: University-wide alert sent through social media (e.g. Facebook, Twitter)

strongly disagree ○—○—○—○—○ strongly agree
1 2 3 4 5

SM6: University-wide alert sent through email (e.g. using official [redacted] emails)

strongly disagree ○—○—○—○—○ strongly agree
1 2 3 4 5

GENERAL FEEDBACK

Please describe other things that may have been relevant to your interaction with this research or scenario

____

____

____

____

THANK YOU FOR TAKING TIME TO PARTICIPATE IN THIS RESEARCH

APPENDIX G: CODE BOOK

	Variable (Code)	*Measurement Statement*	*Measurement Level*	*Possible Values*
	DEMOGRAPHIC DATA			
1.	**Gender (GENDER)**	What is your gender?	Binary	0: Male 1: Female
2.	**Age (AGE)**	What is your age in years?	Ordinal	1: less than 18 years 2: 18 - 25 years 3: 26 - 35 years 4: 36 - 45 years 5: 46 - 55 years 6: above 55 years
3.	**Level of Education (EDUCATION)**	What is the highest level of education you have *completed*?	Ordinal	1: Primary School 2: High School 3: Diploma 4: Undergraduate Degree (Bachelor's) 5: Graduate Degree (Master's) 6: Doctoral Degree (PhD)
4.	**Role (ROLE)**	What is your role at the university?	Nominal	1: Student 2: Faculty/Lecturer 3: Staff 4: Other
5.	**Year first used the internet (YEAR_INTERNET)**	Which year did you first use the internet?	Ordinal	1: before 1991 2: 1991-1995 3: 1996 -2000 4: 2001-2005 5: 2006-2010 6: after 2010
6.	**Hours spent on the internet in a day (HOURS_INTERNET)**	How many hours do you spend on the internet in a day?	Ordinal	1: less than 5 2: 5-10 3: 11-15 4: 16-20 5: 21-24
7.	**Computer Skill (COMP_SKILL)**	How would you rate your computer skills?	Ordinal	1: Low 2: Basic 3: Intermediate 4: Advanced 5: Expert
8.	**Email Load (EL)**	How many emails do you receive in your official USIU-A email account in a day?	Ordinal	1: less than 10 2: 11-20 3: 21-30 4: 31-40 5: 41-50 6: more than 50
9.	**Email Responsiveness**			
	ER1	I *read* all emails I receive in my official USIU-A email account	Ordinal	5 point Likert Scale
	ER2	I *respond* to all emails I need to in my official USIU-A email account	Ordinal	5 point Likert Scale
10.	**Online Services Usage: To what extent do you use the following online services?**			
	OS1	Email	Ordinal	5 point Likert Scale
	OS2	Social Media	Ordinal	5 point Likert Scale
	OS3	Online Shopping	Ordinal	5 point Likert Scale
	OS4	Online Banking	Ordinal	5 point Likert Scale
11.	**Prior Victimization: Have you ever experienced the following online threats in the past?**			
	PV1	Scam	Ordinal	0: No 1: Yes
	PV2	Online Account Hijacking	Ordinal	0: No 1: Yes

	Variable (Code)	*Measurement Statement*	*Measurement Level*	*Possible Values*
	PV3	Identity Theft	Ordinal	0: No 1: Yes
	PV4	Credit/Debit Card Fraud	Ordinal	0: No 1: Yes
	PV5	Malicious software infection	Ordinal	0: No 1: Yes
12.	**Risk Propensity: To what extent do you agree with the following statements about your risk propensity?**			
	RP1	I like taking risks	Ordinal	5 point Likert Scale
	RP2	People say I am a risk taker	Ordinal	5 point Likert Scale
	RP3	I sometimes take risks that could threaten my safety	Ordinal	5 point Likert Scale
MODEL VARIABLES				
1.	***BEHAVIOURAL OUTCOME***			
	QSR_OB1	Did you read this email?	Binary	0: No 1: Yes
	QSR_OB2	Did you click the link labelled “click here” on this email?	Binary	0: No 1: Yes
	DOB_OB2	Observed click behaviour from the website	Binary	0: No 1: Yes
	QSR_OB3	Did you fill in the form presented on the website?	Binary	0: No 1: Yes
	DOB_OB2	Observed form-fill behaviour from the website	Binary	0: No 1: Yes
2.	***THREAT AVOIDANCE***			
	TAV1	My intention was to protect my computer resources	Ordinal	5 point Likert Scale
	TAV2	My intention was to protect my data	Ordinal	5 point Likert Scale
COPING APPRAISAL				
3.	**Response Efficacy**			
	RE1	Enabling security measures would protect users from similar threats	Ordinal	5 point Likert Scale
	RE2	Enabling security measures would prevent users from being deceived by similar threats	Ordinal	5 point Likert Scale
	RE3	Enabling security measures would prevent attackers from successfully launching similar threats	Ordinal	5 point Likert Scale
4.	**Self-Efficacy**			
	SE1	I could learn to protect myself from similar threats without much assistance	Ordinal	5 point Likert Scale
	SE2	It would be easy for me to learn security measures to protect myself from similar threats	Ordinal	5 point Likert Scale
	SE3	I can learn new computer security skills without much difficulty	Ordinal	5 point Likert Scale
5.	**Response Cost**			
	RC1	Taking precautions to prevent such threats would be an inconvenience	Ordinal	5 point Likert Scale
	RC2	Taking precautions to prevent such threats would be time consuming	Ordinal	5 point Likert Scale
	RC3	Taking precautions to prevent such threats would hinder my productivity	Ordinal	5 point Likert Scale
6.	**Perceived Benefit**			
	PB1	Taking precautions to prevent similar attacks would be worthwhile	Ordinal	5 point Likert Scale
	PB2	USIU-A would benefit greatly from protecting its systems from similar attacks	Ordinal	5 point Likert Scale
	PB3	Protecting myself from similar attacks would be beneficial	Ordinal	5 point Likert Scale
ORGANIZATIONAL FACTORS				
7.	**Policies**			
	POL1	I am required to know a lot about USIU-A’s information security policies	Ordinal	5 point Likert Scale

	Variable (Code)	*Measurement Statement*	*Measurement Level*	*Possible Values*
	POL2	I know the regulations outlined in USIU-A's information security policies	Ordinal	5 point Likert Scale
	POL3	USIU-A's information security policies can guide me in handling such threats	Ordinal	5 point Likert Scale
8.	**Technology Controls**			
	TC1	USIU-A has equipped me with technology controls that can detect such threats	Ordinal	5 point Likert Scale
	TC2	USIU-A has equipped me with technology controls that can prevent such threats	Ordinal	5 point Likert Scale
	TC3	USIU-A has equipped me with technology controls that can protect me from such threats	Ordinal	5 point Likert Scale
9.	**Security Education, Training & Awareness**			
	SETA1	USIU-A has made me aware of such threats	Ordinal	5 point Likert Scale
	SETA2	USIU-A has provided me with training on how to handle such threats	Ordinal	5 point Likert Scale
	SETA3	USIU-A has given me sufficient information regarding such threats	Ordinal	5 point Likert Scale
10.	***THREAT DETECTION***			
	TD1	I could tell this was an online attack	Ordinal	5 point Likert Scale
	TD2	I could tell someone was trying to deceive me	Ordinal	5 point Likert Scale
	TD3	I could tell that someone was trying to capture my personal details and password	Ordinal	5 point Likert Scale
THREAT APPRAISAL				
11.	**Perceived Vulnerability**			
	PVUL1	The chances of receiving fraudulent emails are high	Ordinal	5 point Likert Scale
	PVUL2	I am a likely target for online attacks	Ordinal	5 point Likert Scale
	PVUL3	I am likely to encounter various online attacks	Ordinal	5 point Likert Scale
12.	**Perceived Severity: Please rate how bad you think the consequences of the following actions could be on the internet**			
	PS1	Opening a suspicious email	Ordinal	5 point Likert Scale
	PS2	Opening a suspicious attachment	Ordinal	5 point Likert Scale
	PS3	Clicking a suspicious hyperlink	Ordinal	5 point Likert Scale
	PS4	Loading a suspicious website	Ordinal	5 point Likert Scale
	PS5	Filling out personal details on a website	Ordinal	5 point Likert Scale
	PS6	Sharing my USIU-A username and password	Ordinal	5 point Likert Scale
KNOWLEDGE				
13.	**Threat Domain: Please indicate what the following words mean with regards to information security.**			
	KQ	Knowledge Quiz Score	Ratio	0 <= Integer <=6
	KQ1	Phishing	Binary	0: Incorrect; 1: Correct
	KQ2	Social Engineering	Binary	0: Incorrect; 1: Correct
	KQ3	URL	Binary	0: Incorrect; 1: Correct
	KQ4	Certificate	Binary	0: Incorrect; 1: Correct
	KQ5	Spoofing	Binary	0: Incorrect; 1: Correct
	KQ6	Domain	Binary	0: Incorrect; 1: Correct
14.	**Detection Cues**			
	DC1	I know how to reveal hyperlinks hidden behind text to detect such threats	Ordinal	5 point Likert Scale
	DC2	I know how to analyze web addresses to detect such threats	Ordinal	5 point Likert Scale
	DC3	I know how to analyze web certificates to detect such threats	Ordinal	5 point Likert Scale

	Variable (Code)	*Measurement Statement*	*Measurement Level*	*Possible Values*
15.	**Trust Determinants: To what extent did you use the following characteristics or techniques to determine the trustworthiness of the email/website?**			
	DT1	Consistency in logo, colors, look and feel	Ordinal	5 point Likert Scale
	DT2	Grammar and Spelling	Ordinal	5 point Likert Scale
	DT3	Personalized greeting with your names	Ordinal	5 point Likert Scale
	DT4	Content (e.g. reasonableness of the explanation in email and website content)	Ordinal	5 point Likert Scale
	DT5	Context (e.g. it was expected in the prevailing circumstances)	Ordinal	5 point Likert Scale
	DT6	Email address of the sender	Ordinal	5 point Likert Scale
	DT7	Contacting the USIU-A ICT helpdesk	Ordinal	5 point Likert Scale
	DT8	Asking someone (e.g. colleague, friend)	Ordinal	5 point Likert Scale
	DT9	Web address and hyperlink evaluation	Ordinal	5 point Likert Scale
	DT10	Website encryption or padlock icon	Ordinal	5 point Likert Scale
	DT11	Website certificate	Ordinal	5 point Likert Scale
	DT12	Domain registration information (e.g. from whois)	Ordinal	5 point Likert Scale
	DT13	Security tool information (e.g. anti-phishing tool integrated in email/browser)	Ordinal	5 point Likert Scale
16.	***ELABORATION***			
	ELAB1	I made conscious effort to evaluate the email/website	Ordinal	5 point Likert Scale
	ELAB2	I took time to evaluate the email/website	Ordinal	5 point Likert Scale
	ELAB3	I carefully evaluated the email/website	Ordinal	5 point Likert Scale
ATTACK FACTORS				
17.	**Quality of Argument**			
	QA1	I carefully scrutinized the email message before responding	Ordinal	5 point Likert Scale
	QA2	I reasoned through the explanation given in the email before responding	Ordinal	5 point Likert Scale
	QA3	I examined the reasons given in the email before responding	Ordinal	5 point Likert Scale
18.	**Persuasive Cues: Please rate to which extent the following components of the email/website influenced your response**			
	PC1	Source credibility (i.e. ICT administrator)	Ordinal	5 point Likert Scale
	PC2	Personalized Greeting	Ordinal	5 point Likert Scale
	PC3	Offer to extend your mail quota	Ordinal	5 point Likert Scale
	PC4	Warning that your email service would be discontinued	Ordinal	5 point Likert Scale
	PC5	Urgency to respond within 24 hours	Ordinal	5 point Likert Scale
	PC6	Resemblance to other USIU-A emails	Ordinal	5 point Likert Scale
	PC7	Resemblance to other USIU-A websites	Ordinal	5 point Likert Scale
MOTIVATED TO PROCESS				
19.	**Involvement**			
	INV1	The email seemed very relevant to me	Ordinal	5 point Likert Scale
	INV2	The email seemed very important to my work/studies	Ordinal	5 point Likert Scale
	INV3	The email seemed very applicable to my current situation	Ordinal	5 point Likert Scale
20.	**Responsible**			
	RES1	I am answerable to communications I receive on my USIU-A email account	Ordinal	5 point Likert Scale
	RES2	I am in control of the day-to-day operation of my USIU-A email account	Ordinal	5 point Likert Scale
	RES3	I consider myself responsible for my USIU-A email account	Ordinal	5 point Likert Scale

	Variable (Code)	*Measurement Statement*	*Measurement Level*	*Possible Values*
ABILITY TO PROCESS				
21.	**Distraction**			
	DIST1	There is usually a lot of activity going on around me when reading and responding to emails	Ordinal	5 point Likert Scale
	DIST2	I usually multi-task when reading and responding to emails	Ordinal	5 point Likert Scale
	DIST3	I tend to be distracted when reading and responding to emails	Ordinal	5 point Likert Scale
22.	**Emotions**			
	EM1	Reading the email invoked an emotion in me (e.g. fear, anxiety)	Ordinal	5 point Likert Scale
	EM2	I responded to this email so that I would not get into trouble	Ordinal	5 point Likert Scale
	EM3	I would have felt guilty for not responding to the email	Ordinal	5 point Likert Scale
23.	**Pressure**			
	PRES1	I am usually under pressure to move on to other tasks when reading and responding to emails	Ordinal	5 point Likert Scale
	PRES2	I usually have a sense of urgency when reading and responding to emails	Ordinal	5 point Likert Scale
	PRES3	I tend to rush through my emails	Ordinal	5 point Likert Scale

APPENDIX H: JOURNAL PUBLICATIONS

Computers in Human Behavior 94 (2019) 154–175

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com/locate/comphumbeh

Full length article

A new approach to modelling the effects of cognitive processing and threat detection on phishing susceptibility

Paula M.W. Musuva[a,*], Katherine W. Getao (PhD, EBS)[b], Christopher K. Chepken (PhD)[c]

[a] United States International University-Africa, P.O. Box 14634, 00800, Nairobi, Kenya
[b] ICT Authority, P.O. Box 27150 - 00100, Nairobi, Kenya
[c] University of Nairobi, School of Computing and Informatics, P.O. BOX 30197, Nairobi, Kenya

ARTICLE INFO

Keywords:
Social engineering
Phishing
Elaboration
Threat detection
Elaboration likelihood model
Structural equation modelling

ABSTRACT

Phishing threats are real and are ever increasing in their reach and devastating effects. This study delves into the role of cognitive processing in detecting and curtailing phishing attacks. The proposed model is grounded on the Elaboration Likelihood Model and is tested empirically using data from 192 cases. Data was collected through direct observations of phishing susceptibility and self-reported questionnaires after staging a phishing attack targeting a university population in Nairobi, Kenya. The model was found to have excellent fit and was able to account for 50.8% of a person's cognitive processing of a phishing attack, 69.5% of their ability to detect phishing threats and could predict 28% of their actual phishing susceptibility. Analysis was done to test 25 hypothesis, and to examine the mediating effects of cognitive processing and threat detection. In addition, multi-group moderation analysis was done to examine if the model was invariant based on the level of knowledge. Results indicate that threat detection has the strongest effect in reducing phishing susceptibility. Threat detection was found to be what explains why people who expend cognitive effort processing phishing communication are less likely to fall for phishing threats.

1. Introduction

Information security attacks are constantly making headlines all over the world. Many of these cases are facilitated by valid users of information systems; whether knowingly or accidentally. This phenomenon has been termed as the insider threat (CERT, 2013). Insiders are often deceived by malicious outsiders to undertake actions that compromise their systems. This is because hackers have for a long time known this to be the Achilles heel of otherwise highly secured systems (Mitnick & Simon, 2002). A popular technique used to deceive insiders is phishing.

Phishing is the use of technical mediums (such as email, websites, chat or text messages) to deceive users into divulging sensitive and confidential information; such as: identity profiles, usernames and passwords (APWG, 2018).

The term ‘phishing’ is derived from the word ‘fishing’ with the intended metaphorical symbolism. The deceptive communication is set up as the bait with the intention of fooling a user into thinking it is a legitimate request for information. The technical mediums, such as, emails or websites are like the fishing rods used to reel in the catch that is unlucky enough to fall for the bait. The catch is the confidential information that eventually provides access and use of a protected information system.

Phishing has been used for decades. James (2005) dates it as far back as 1995. Despite it having a long history of practice, reported cases are still on the rise. The Anti-Phishing Working Group report (APWG, 2017) puts the percentage increase of phishing attacks reported in the fourth quarter of 2016 at 65% compared to those reported in 2015. A previous report for the first quarter of 2016 (APWG, 2016) reveals a 250% increase in the number of unique phishing websites over a period of just 6 months. Moreover, a trend analysis reveals a 5753% increase of phishing attacks in a 12 year period since 2004. Another organization that monitors Phishing activity, PhishTank, recorded the presence of 4.5 million phishing sites in October 2016 (PhishTank, 2016).

The financial costs of phishing attacks are also striking. Cyveillance (2015) estimates annual losses of 5.9 billion US dollars due to phishing attacks. Another news report in 2016 (Barth, 2016; BBC News, 2016) unveiled a group of criminals led by a Nigerian man that scammed businesses and individuals of 60 million US dollars using phishing malware and email-based scams.

The Anonymous hacktivist group has also used phishing to compromise government systems in the African continent. A case in point is

[*] Corresponding author.
E-mail address: pmusuva@usiu.ac.ke (P.M.W. Musuva).

https://doi.org/10.1016/j.chb.2018.12.036
Received 20 August 2018; Received in revised form 7 December 2018; Accepted 21 December 2018
Available online 04 January 2019

the compromise of Kenya's Ministry of Foreign affairs; where 1 Tera Byte (TB) of its data was posted on the dark web (Cimpanu, 2016; Obulutsa, 2016; Waqas, 2016).

Phishing has also been used for political purposes. Investigative reports by Fire Eye (2017) on Russia's involvement in the United States 2016 presidential elections shows the use of spear-phishing emails to compromise key staff in the Democratic Party.

The phishing threat is present and is ever increasing in its reach and devastating effects. This study examines the effect a user's thought processing has on detecting and curtailing phishing attacks. It explores the role that cognitive evaluation of phishing messages received by a user has on their ability to identify an attack and stop it.

2. Background

Various studies have tried to understand why people succumb to phishing attacks.

Dhamija, Tygar, and Hearst (2006) found that phishing works because people use ineffective criteria to assess the credibility of phishing communication. They found that 36% of the study participants used the domain name address and content on a phishing page (such as; layout, logos, graphics, links, language and accuracy of information) to determine legitimacy of phishing messages. These were poor indicators of credibility. Only 9% examined for HTTPS encrypted sessions and only 9% checked the certificate; yet these were better indicators of credibility. Disturbingly, they found that 68% of the participants still proceeded to access websites for which their browsers raised warnings. In fact, 90% of the participants in their study were fooled by well-designed phishing websites.

Downs, Holbrook, and Cranor (2006, 2007) further examined these issues in their study and found that there were three predictors of behaviour: knowledge, perception of trustworthiness and perception of negative consequences. On knowledge, they found that participants who correctly answered knowledge questions relating to phishing were significantly less likely to fall for phishing attacks. This was very specific to phishing knowledge because the finding did not apply to those who had correct general knowledge on unrelated computer risks (such as; viruses). On the perception of trustworthiness, they found that a person's judgement of trustworthiness was a significant predictor of behaviour. A person has to assess phishing communication to be untrustworthy and phishing sites as insecure for them not to click links or enter information. On perception of negative consequences, they found that one had to associate a likelihood of negative consequence in order to avoid insecure actions. People would not give their social security numbers and sensitive information when they feared negative consequences.

Jakobsson, Tsow, Shah, Blevis, and Lim (2007) conducted a qualitative study of phishing to delve deeper into what instills trust. They found that people trusted phishing messages that were well written, had proper spelling and grammar. They also trusted messages which were personalized to them; for example, by mentioning their name and contact information. Also the presence of high-resolution padlock icons, legal disclaimers in fine print and third party endorsements (such as the presence of Verisign badge) made many trust messages. If the phishing message gave a number for someone to call, the participants said that they would trust the message more – even if they admitted that they would not actually call the number. On the contrary, they did not trust messages that had spelling mistakes, were not signed off by a person and had syntactically strange addresses.

Kumaraguru, Sheng, Acquisti, Cranor, and Hong (2007) were able to further study trust determinants and classified them in seven categories. The first category examined the use of design and content. They found that 42% of users used the design and content of the phishing message and ascribed credibility when the look and feel was professional, links were working, and contact information had been provided. The second category was the use of the URL and domain name. They found that 31% judged domain names that had numbers or that were poorly formed to be suspicious. The third category was the information requested by the phishers. They found that 19% of users judged communication to be suspicious if it requested a lot of information and especially if that information was sensitive. The fourth category was the use of search engines. They found that 16% of the participants went onto search engines and checked search results to determine if the sites were legitimate. A similar technique makes use of security toolbars that give a rating for websites. The fifth category was consistency. They found that 16% used consistency in logos and colors from their familiarity with legitimate sources to judge credibility. The sixth category was prior knowledge. They found that 6% of users used their experience with similar attempts to identify an attack. Finally, the seventh category was the use of security indicators. They found that only 3% examined security indicators such as HTTPS padlocks or certificates to examine legitimacy.

These studies have shown that people fall for phishing attacks due to lack of knowledge or the use of incorrect criteria in determining the trustworthiness of phishing messages. This has prompted many to recommend user education as a countermeasure. Consequently, studies have examined how this training should be effectively conducted (Kumaraguru et al., 2009; Kumaraguru, Sheng, Acquisti, Cranor, & Hong, 2008; Kumaraguru, Rhee, Acquisti, et al., 2007; Kumaraguru, Rhee, Sheng, et al., 2007; Kumaraguru, Sheng, et al., 2007; Sheng et al., 2007). These studies have found that embedding phishing training in real-world scenarios is the most effective way of conducting user education. Embedded training involves presenting users with training materials immediately they click or interact with phishing artifacts. This requires organizations to integrate phishing tests in normal day-to-day communication. Organizations also need a training system that immediately gives users feedback on insecure behaviour and points out lessons they can learn from their interactions with phishing content.

Other studies have taken a different approach by examining the persuasive and deceptive techniques used by attackers. Jagatic, Johnson, Jakobsson, and Menczer (2007) demonstrated the use of data gleaned from crawling social networking sites to craft spoofed messages appearing to come from social network friends in order to increase veracity of phishing attacks. Luo, Zhang, Burd, & Seazzu, (2013) showed that attackers can craft messages with high argument quality, source credibility and conformity to genres of accepted communication, in order to increase potency of phishing attacks. Workman (2007, 2008a, 2008b) examined the use of commitment, likeability, trust, fear, authority and scarcity in manipulating people to fall for social engineering.

Unfortunately, few of these studies have been grounded in theory. In fact, Tetri & Vuorinen (2013) point out that in their review of 40 studies on social engineering, only two had explicit underlying theories in their research. Luo et al. (2013) also point out that there is little research that is grounded in theory that explains why people fall victim to phishing attacks. Little research has gone into systematically identifying and analyzing the factors that make phishing attacks successful. Wang, Herath, Chen, Vishwanath, & Rao (2012) call for more research that can examine the theoretical foundation that explain how people process phishing communication with empirical evidence.

This study is a response to this gap. It presents an empirical study, grounded in theory, used to analyze cognitive processing and threat detection when determining why individuals fall for phishing attacks. This study particularly pays attention to peoples' thought processing mechanisms when under attack and also their ability to detect threats. Theories that explore this causal mechanisms are presented and tested to better understand phishing susceptibility.

The analysis of the cognitive processing of phishing messages is an area that has recently drawn attention in phishing research. This is because phishing takes advantage of people's weaknesses in processing communication as opposed to taking advantage of technology loopholes. The aim of the attacker is to evade threat detection and to encourage insecure responses to phishing communication.

Luo et al. (2013) conducted a field study that examined a real spear phishing campaign that targeted 105 insiders at a public university in Southwest USA. They grounded their research on the Heuristic-Systematic Model in order to provide a theoretical model that systematically describes the cognitive model associated with phishing susceptibility. They found that 36% of the targeted users clicked the email phishing link while 15% of them submitted their login credentials on the phishing login page.

Wang et al. (2012) did research to determine how people process phishing by specifically examining how they pay attention to visceral (deceptive) cues and also phishing indicators when making a decision to respond to phishing messages. Their model was able to explain 16% of an individual's likelihood to respond to phishing emails. It was also able to explain 11% of the cognitive processing effort that subsequently affects phishing susceptibility.

Vishwanath, Herath, Chen, Wang, & Rao (2011) conducted a study with the aim of presenting a single comprehensive and integrated model that explains phishing susceptibility. Their model sought to explain combined effects of various factors by simultaneously examining how individuals process phishing, which aspects they paid attention to and how individual-based factors affected cognitive evaluation and subsequently their susceptibility to phishing. Their model was able to account for 46% of the variance in a person's likelihood to respond to phishing and 22% of elaboration and 22% of attention paid to different elements of the email (namely, source, grammar/spelling, subject and urgency).

A summary of previous studies that examined how cognitive processing influences phishing susceptibility are as outlined in Table 1.

As highlighted in Table 1, the theories that have been advanced in this area of research are the: Interpersonal Deception Theory (IDT), Theory of Deception (ToD), Heuristic Systematic Model (HSM) and Elaboration Likelihood Model (ELM).

The Interpersonal Deception Theory (IDT) by Buller and Burgoon (1996) has been used in a previous study by Vishwanath et al. (2011) as a theory of interest in understanding phishing attacks. It analyzes deception when it takes place in interactive contexts; which are mostly face-to-face encounters. The sender (deceiver) and the receiver (deceived) gauge each other's responses and adapt their behaviour during the deception process. Therefore, the sender is able to strategically alter their message based on the responses they observe from the receiver (even if they are non-verbal) in order to carry out successful deception.

The Interpersonal Deception Theory may be useful when examining cases of phishing delivered through active inter-personal engagement between an attacker and an insider; for example, when handled through a phone conversation (a technique referred to as vishing). The key element here is the ability of the attacker to evaluate the responses from the insider in order to adapt their deception. However, in cases where there is no immediate feedback and interactive engagement, this theory may not be suitable.

Another theory of interest is the Theory of Deception (ToD) advanced by Johnson, Grazioli, Jamal, and Zualkernan (1992); Johnson, Grazioli, Jamal, and Berryman (2001); and Grazioli (2004). It has been commonly applied to understand how consumers of information detect deceptive communication. It is largely similar and consistent with the Interpersonal Deception Theory which focuses on an individual's information processing during deception.

However, the Theory of Deception (ToD) differs from the Interpersonal Deception Theory in 3 areas (Grazioli, 2004; Johnson et al., 2001; Vishwanath et al., 2011). Firstly, the Interpersonal Deception Theory focuses mostly on communication and social psychology; while the Theory of Deception has found use in more disciplines and business contexts. Of particular interest are studies in information and communication technology (Grazioli, 2004; Grazioli & Jarvenpaa, 2001; Vishwanath et al., 2011) that examine online deceptions occurring on the internet. Secondly, the Theory of Deception covers deceptions that have lower interactivity between the deceiver and target. It focuses on those that involve the evaluation of content as opposed to the high interactivity that the Interpersonal Deception Theory addresses. Thirdly, the Theory of Deception does not focus on the interplay between the deceiver and the target. Rather, it focuses on the cognitive processing that occurs in the target when they are interacting with the deceptive communication. It examines the mental processing by the target and their ability to reason through the deception while emphasizing their need to have sufficient and competent knowledge regarding the deception. Particularly of the deception techniques used and also the cues that can be used to detect the deception (Vishwanath et al., 2011).

This theory fits very well in the cases of social engineering because it systematically guides the evaluation of the cognitive processes undertaken by insiders to identify gaps that lead to successful attacks. It also emphasizes the need to evaluate the insider's domain specific knowledge and their understanding of detection cues. Various studies have shown that these are key to understanding susceptibility to phishing (Dhamija et al., 2006; Downs, Holbrook, & Cranor, 2007, 2006; Grazioli, 2004; Kumaraguru, Sheng, et al., 2007; Vishwanath et al., 2011).

One weakness in the Theory of Deception is its inability to distinguish different types of cues that could be evaluated when detecting deception. For example, some studies have shown that if insiders focus on persuasive cues they are more likely to fall for deception than if they focus on threat detection cues or even the quality of argument expressed in the phishing communication (Luo et al., 2013; Vishwanath et al., 2011). Another weakness is that the Theory of Deception does not address the influence that emotional factors have on the detection of deception. It only approaches deception from a rational thinking perspective (Grazioli, 2004).

In order to address these deficiencies other theories are proposed to examine the case of phishing susceptibility; namely the Heuristic Systematic Model (Chaiken, 1980) and the Elaboration Likelihood Model (Petty & Cacioppo, 1986). Both the Elaboration Likelihood Model (ELM) by Petty & Cacioppo (1986) and Heuristic Systematic Model (HSM) by Chaiken (1980) propose two cognitive processing modes during the evaluation of persuasive communication. These dual-processing theories provide a fuller explanation compared to one-process approaches advanced by Theory of Deception and others such as Cognitive Dissonance Theory (Festinger, 1957) and Reactance Theory (Brehm, 1966) that have also been evaluated in social engineering research (Workman, 2007; 2008b).

The first cognitive processing mode is termed as “Central” in ELM or “Systematic” in HSM. It is characterized by a person's careful reasoned evaluation of the issue-relevant arguments presented by the persuasive communication.

The second cognitive processing mode is described as “Peripheral” in ELM or “Heuristic” in HSM and is characterized by low cognitive processing of the issue-relevant arguments. Instead, reliance is placed on simple peripheral (persuasive) cues to make judgment. Peripheral cues are used to short-circuit logical reasoning and often invoke quick responses that are not well thought out.

The Elaboration Likelihood Model (ELM) and Heuristic Systematic Model (HSM) are very similar. Firstly, in the description of the cognitive evaluation process in dual modes. Secondly, they both assume that people have a desire to hold onto what they judge to be the correct attitudes or judgment for a given scenario. The correctness could be determined by their evaluation of the arguments presented in the message but also their reliance on certain persuasive cues. Thirdly, both suggest that engagement in the higher cognitive effort (Central in ELM or Systematic in HSM) is driven by the processing of issue-relevant arguments. They also both agree that long-lasting attitudes and behaviour changes are affected by this higher cognitive effort.

Although ELM and HSM are largely similar, they have a few important differences as pointed out by the model developers (Eagly &

Table 1
Summary of previous studies on cognitive processing of phishing.

Study	Description	Analysis
Luo et al. (2013) Investigating Phishing Victimization with the Heuristic-Systematic Model: A Theoretical Framework and an Exploration	● Field study ● Qualitative review ● Actual spear phishing attack that targeted 105 faculty and staff at a public university located in Southwest US **Theory:** Heuristic-Systematic Model **Variables:** ● Argument Quality ● Source Credibility ● Genre Conformity ● Need for Cognition ● Time Pressure ● Pre-texting ● Less Damage ● DV: Victimization	Key findings: ● Susceptibility: 36% clicked link and 15% gave login credentials ● Factors that have significant effect on phishing susceptibility are: argument quality, source credibility, genre conformity, pretexting, less damage. ● Factors not conclusively tested: need for cognition, time pressure Key limitations: ● Explorative study that did not exhaustively test the proposed set of Hypothesis ● Hypothesis not empirically tested ● Results discussed qualitatively without supporting quantitative depth
Wang et al. (2012) Phishing Susceptibility: An Investigation Into the Processing of a Targeted Spear Phishing Email	● Field study ● Online Questionnaire Survey ● Involved 321 members of a Northeast USA public university who had been targeted by a spear phishing attack ● 267 good cases used for data analysis **Theory:** Theory of Deception **Variables:** ● Attention to Visceral Triggers - Title of email message - Urgency ● Attention to Phishing Deception Indicators - Grammar - Sender's Address ● Cognitive Effort ● Message Involvement ● Scam Knowledge ● Demographic Factors: - Gender - Age - Knowledge of Emails from the organization ● **DV:** Likelihood to Respond	Key findings: ● Model was able to account for 16.4% of the variance in a person's likelihood to respond to phishing and 11.9% of cognitive effort ● Attention to visceral triggers significantly reduced cognitive processing and significantly increased the likelihood to respond to phishing ● Attention to phishing deception indicators marginally reduced cognitive processing but significantly reduced the likelihood to respond to phishing ● Cognitive processing effort did not significantly reduce the likelihood to respond to phishing ● Message involvement significantly increased cognitive processing ● Knowledge of email-based scams significantly increased attention to phishing deception indicators ● Knowledge of email-based scams marginally reduced likelihood to respond to phishing ● Knowledge of email-based scams weakened (moderated) effect of attention to visceral triggers on likelihood to respond to phishing ● Knowledge of email-based scams strengthened (moderated) effect of attention to phishing deception indicators on likelihood to respond to phishing Key limitations: ● Used one email sample that does not address many characteristics of phishing ● Used single-item measure for construct measuring knowledge of email-based scams ● Cognitive processing effort measures not fully developed ● Data collected using single-round survey. Multiple methods of data collection could be triangulated ● Self-reported questionnaire measures may not be as reliable as observed behaviour ● Convenience sample of university students is not good representation of general population. Therefore results are not highly generalizable
Vishwanath et al. (2011) Why do people get phished? Testing individual differences in phishing vulnerability within an integrated, information processing model	● Field study ● Online Questionnaire Survey ● Received 325 responses whereby 4 were overlapping thus 321 usable responses ● Used split-half method of model testing. Divided data into two samples of 161 and 160 for initial and confirmatory testing respectively ● Examined 2 real phishing attacks sent to users at a Northeast USA university **Theories:** ● Interpersonal Deception Theory (IDT) ● Theory of Deception ● Elaboration Likelihood Model **Variables:** ● Involvement ● Computer Self-Efficacy ● Domain Specific Knowledge ● Email Load ● Attention to sender source ● Attention to grammar and spelling ● Attention to Urgency ● Elaboration	Key findings: ● Final model was able to account for 46% of variance in a person's likelihood to respond to phishing, 22% of elaboration and 22% of attention ● When Attention and Elaboration mediated the influence of involvement, the model accounted for 20% of the variance ● Attention to email's source, grammar/spelling significantly reduced likelihood to respond to phishing ● Attention to urgency cues and subject line significantly increased likelihood to respond to phishing ● Attention to urgency cues significantly reduced elaboration. Attention to other elements (source, grammar/spelling, subject) did not have significant impact ● Elaboration reduced likelihood to respond to phishing but effect was not significant ● Involvement significantly increased attention to urgency cues. It did not have significant impact to other elements (source, grammar/spelling, subject) ● Involvement significantly increased Elaboration

(*continued on next page*)

Table 1 (*continued*)

Study	Description	Analysis
	● DV: Likelihood to respond	● Involvement significantly increased Likelihood to Respond ● Email load did not significantly reduce attention given to various elements (source, grammar/spelling, urgency, subject) of the email ● Email load significantly increased likelihood to respond to phishing ● Domain specific knowledge partially increased Elaboration ● Computer Self-efficacy did not significantly increase Elaboration Key limitations: ● Self-reported questionnaire measures may not be as reliable as observed behaviour ● Required users to recall past events which may have affected quality of data

Chaiken, 1993; Petty, 1994; Petty & Wegener, 1998). HSM posits that heuristic rules are knowledge structures that are kept in memory and accessed by an individual when they are evaluating a persuasive communication. In addition, HSM presents the concept of the "sufficiency threshold" whereby an individual only engages in evaluating a message until the sufficiency threshold is reached. When some initial heuristic processing does not meet the threshold then systematic processing is engaged. In contrast, ELM recognizes heuristic processing as just one of a number of possible peripheral route processes. In ELM there is a trade-off (negative relationship) between central and peripheral processing thereby giving a distinction for underlying attitude-forming processes as opposed to HSM in which both modes augment each other.

Due to these differences, and also the considered view that ELM evaluates a multi-dimensional space of the source, message, recipient and contextual factors (Petty & Wegener, 1999); it is proposed that ELM be used in this research. Additionally, ELM has been explored more widely in information system research; such as studies by: Wang et al. (2012) and Vishwanath et al. (2011) on factors that lead to phishing susceptibility; Angst and Agarwal (2009) on the acceptance of Electronic Health Record systems; Workman (2007, 2008b) on phishing and pretext social engineering; LaRose, Rifon, and Enbody (2008) on improving users' online security behaviour; Bhattacherjee and Sanford (2006) on accepting new Information Technologies; and Johnson et al. (2001) on in the detection of financial fraud.

A key construct in ELM is "Elaboration" which describes the mental effort an individual engages when evaluating persuasive communication. High elaboration means the individual is engaged in high levels of objective information processing and is associated with the central route of information processing. Low elaboration means the individual is engaged in low levels of biased information processing which tends toward subjective reasoning associated with the peripheral route. This a key factor to deception detection. In fact, Vishwanath et al. (2011) demonstrated that successful phishing attempts are mostly characterized by low elaboration. Luo et al. (2013) point out that an attacker's aim is to generate phishing communication that discourages objective, systematic processing but encourages attention to deceptive peripheral cues that result in quick and incorrect decisions.

The deceptive peripheral cues identified in previous studies on phishing susceptibility include: spelling and grammar; professional look and feel; genre conformity; security padlock icons; endorsements; spoofed or falsified source credibility; hiding the deception behind text or images; pretexting; urgency and time pressure (Downs et al., 2006; Jakobsson et al., 2007; Kumaraguru, Sheng, et al., 2007; Luo et al., 2013; Vishwanath et al., 2011).

It is also important to emphasize that for an insider's thought processing to be successful in identifying deception, they should be knowledgeable on both the threat domain and the deception cues. Various studies have demonstrated this (Dhamija et al., 2006; Dodge, Carver, & Ferguson, 2007; Downs et al., 2007, 2006; Jakobsson et al., 2007; Vishwanath et al., 2011) and have shown that the more knowledgeable insiders are on the threat domain and detection cues, the less likely they are to succumb to unintentional threats. These studies have also pointed out that this knowledge could be obtained from training and awareness activities but also from an insider's past exposure to a similar threat.

The Elaboration Likelihood Model is illustrated in Fig. 1.

The concept of "Elaboration" is similar to the concept of "Activation" that is advanced by the Theory of Deception. However, unlike the Theory of Deception, ELM provides differentiation of the information processed into the categories of issue-relevant arguments and peripheral cues and examines the effect that paying attention to these different components has in detection of deception. This enables us to elucidate the different components of deception and examine their effect on the detection of deception.

Another key contribution that distinguishes ELM from other theories is that it seeks to understand what would make the individual (1) motivated to process the persuasive communication presented to them and also (2) what would interfere with their ability to process it objectively. It posits that people will be motivated to process persuasive communication if they feel involved in or responsible for the matter presented. The more a person is motivated to process, the higher their levels of elaboration. On the other hand, their ability to process is hindered by factors such as distraction, emotions or pressure.

A summary of the different theoretical models that have been considered in this study and their key differences is summarized in Table 2. The table also gives a summary of the reasons for the selection of the Elaboration Likelihood Model (ELM) as the theoretical foundation underpinning this research.

3. Conceptual model

This study outlines a conceptual model grounded on the Elaboration Likelihood Model (ELM) as justified the previous section. The conceptual model presented in this study is illustrated in Fig. 2. In this model, Threat Detection and Elaboration are proposed as mediators in the relationship that the various antecedent constructs have on Phishing Susceptibility.

The model constructs are discussed hereafter.

3.1. Phishing susceptibility

The outcome variable under study is Phishing Susceptibility. Two sets of actions are examined. The first set includes actions that are regarded as secure responses to phishing; that is, not clicking on phishing

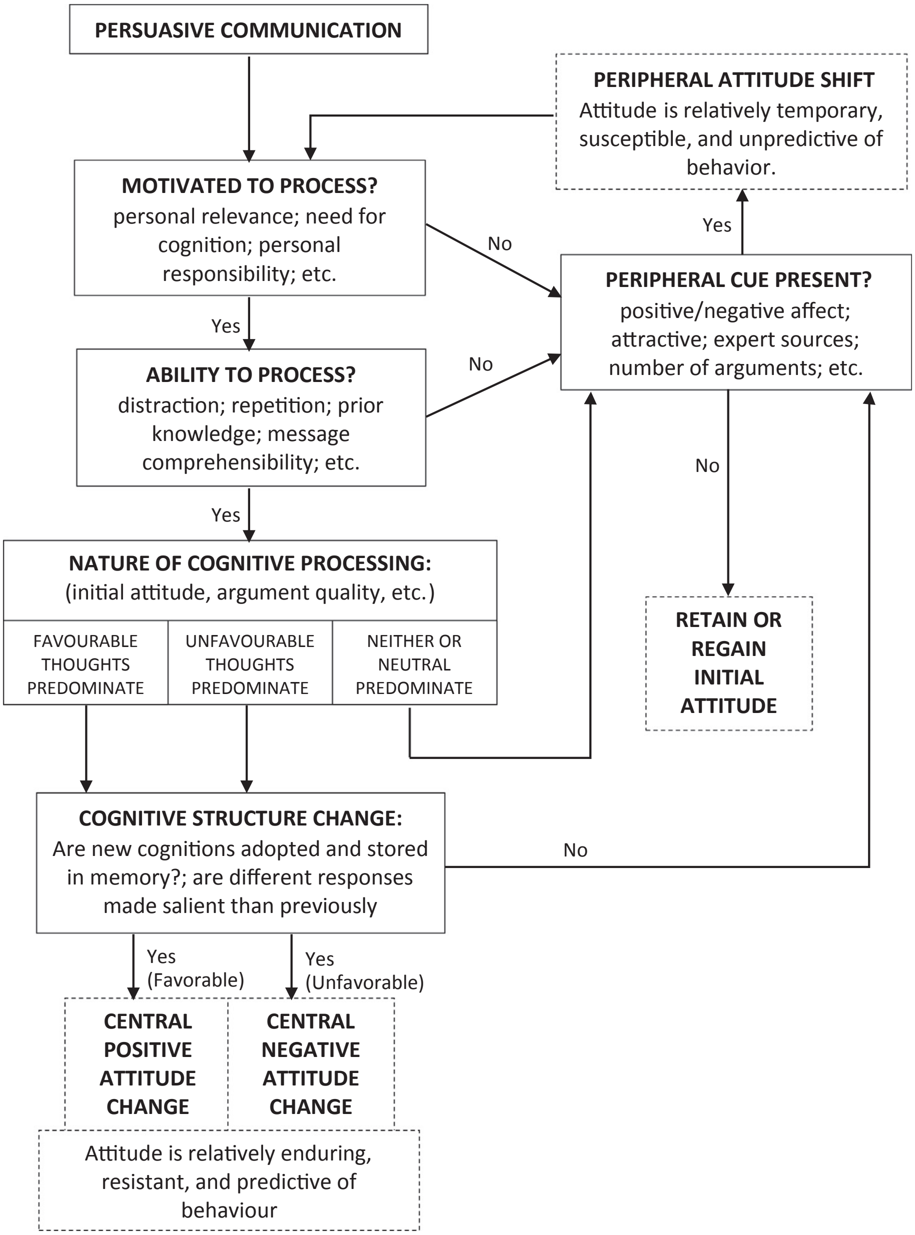

PERSUASIVE COMMUNICATION
PERIPHERAL ATTITUDE SHIFT
Attitude is relatively temporary, susceptible, and unpredictive of behavior.
MOTIVATED TO PROCESS?
personal relevance; need for cognition; personal responsibility; etc.
No
Yes
PERIPHERAL CUE PRESENT?
positive/negative affect; attractive; expert sources; number of arguments; etc.
Yes
ABILITY TO PROCESS?
distraction; repetition; prior knowledge; message comprehensibility; etc.
No
No
Yes
NATURE OF COGNITIVE PROCESSING:
(initial attitude, argument quality, etc.)
FAVOURABLE THOUGHTS PREDOMINATE
UNFAVOURABLE THOUGHTS PREDOMINATE
NEITHER OR NEUTRAL PREDOMINATE
RETAIN OR REGAIN INITIAL ATTITUDE
COGNITIVE STRUCTURE CHANGE:
Are new cognitions adopted and stored in memory?; are different responses made salient than previously
No
Yes (Favorable)
Yes (Unfavorable)
CENTRAL POSITIVE ATTITUDE CHANGE
CENTRAL NEGATIVE ATTITUDE CHANGE
Attitude is relatively enduring, resistant, and predictive of behaviour

Fig. 1. Elaboration Likelihood Model (ELM) (Petty & Cacioppo, 1986).

links or filling-in phishing forms. The second set includes actions that are considered insecure; specifically, clicking of phishing links or filling phishing forms. The outcome variable is measured as a dichotomous categorical variable with a '1' value indicating that an individual was susceptible to phishing and a '0' value indicating that the individual was not susceptible to phishing.

Table 2
Justification of Theoretical Foundation and Construct Selection.

Theories Considered	Key Differences	Justification of Theory Selection	Constructs from ELM
1. Interpersonal Deception Theory	● Suited for interactive and inter-personal contexts	**Elaboration Likelihood Model (ELM) is used because it is:** ● Dual-processing unlike Interpersonal Deception Theory or Theory of Deception ● Multi-dimensional, particularly good when examining both the attacker and target i.e. (1) attack factors and (2) insider cognitive processing ● Empirically tested in various information systems studies and demonstrated to be appropriate for Unintentional Insider Threat Research by Vishwanath et al. (2011) and Workman (2007, 2008b)	● Elaboration ● Quality of Argument ● Persuasive Cues ● Motivated to Process - Involvement - Responsible ● Ability to Process - Distraction - Emotions - Pressure
2. Theory of Deception	● Inability to distinguish different types of cues when analyzing deception		
3. Heuristic Systematic Model	● Dual-processing ● Systematic processing only engaged when sufficiency threshold in heuristic processing is met.		
4. Elaboration Likelihood Model	● Dual-processing ● Multi-dimensional view (attacker, target, message and context)		

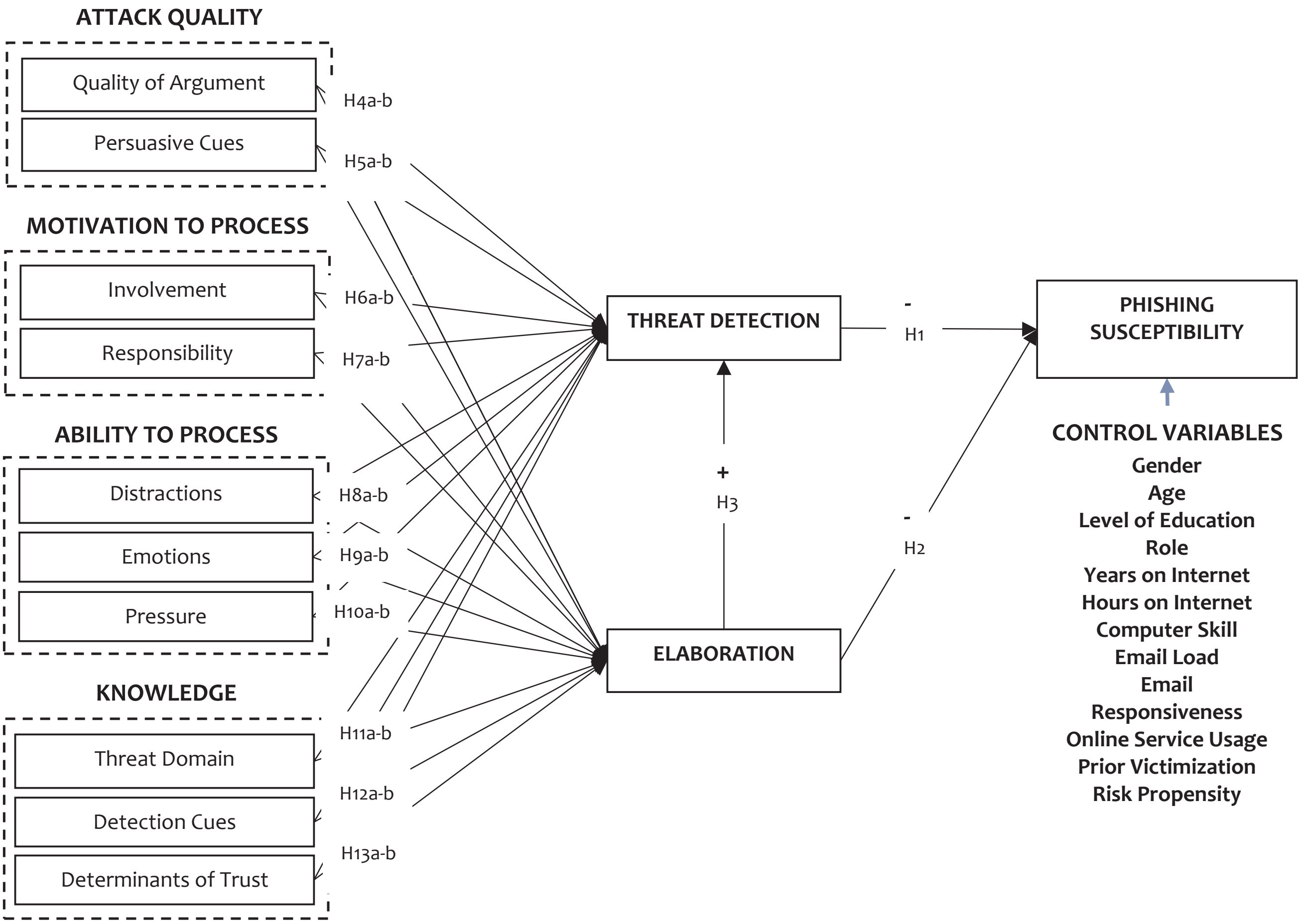


Fig. 2. Conceptual Model on Phishing Susceptibility.

3.2. Threat detection

Threat Detection is one of the antecedent factors that will be studied. Threat Detection is the extent to which a person who is targeted will be able to correctly perceive the phishing attack. This construct has been previously studied by Arachchilage and Love (2013) and Liang and Xue (2009, 2010) using the term Perceived Threat. The term 'Threat Detection' has been preferred over 'Perceived Threat' because it is consistent with the broader information security concept of intrusion detection. In addition, tools developed to counter phishing employ various techniques to assist end users detect possible attacks. This same concept can therefore be studied where anti-phishing tools are used to support users. Arachchilage and Love (2013) and Liang and Xue (2009, 2010) found that threat detection is what prompts an individual to act in a secure manner. Therefore, this study hypothesizes that:

Hypothesis 1. *Threat Detection* has a negative effect on *Phishing Susceptibility*.

3.3. Elaboration

Elaboration is the next construct in the model and is borrowed from the Elaboration Likelihood Model by Petty & Cacioppo (1986). Elaboration is the extent to which a person cognitively evaluates a phishing message by processing the issue-relevant arguments as opposed to dismissively glancing at the message because of its peripheral (or persuasive) cues. High Elaboration takes place when more cognitive effort is given to scrutinize the issue-relevant arguments presented regarding an issue. This seeks to objectively sift truth from fallacy. Low Elaboration takes place when less cognitive effort is dedicated in evaluating a persuasive message. Instead, the validity of a message is judged subjectively based on persuasive cues, such as its professional look and feel.

Elaboration has been found to have an effect on whether users detect threats or not. Wang et al. (2012) shows that susceptibility to phishing emails is dependent on the cognitive effort expended in processing phishing emails. They showed that the likelihood to respond to phishing emails increases with low levels of elaboration. Conversely, with high levels of elaboration users are unlikely to fall for a phishing attempt. Similarly, Vishwanath et al. (2011) in their study show a negative relationship between Elaboration and susceptibility to phishing. Therefore, this study hypothesizes that:

Hypothesis 2. *Elaboration* has a negative effect on *Phishing Susceptibility*.

Unlike the previous studies by Wang et al. (2012) and Vishwanath et al. (2011), this study also examines the relationship between Elaboration and Threat Detection whereby Threat Detection may have a mediating relationship between Elaboration and Phishing Susceptibility. Therefore, this study hypothesizes that:

Hypothesis 3a. *Elaboration* has a positive effect on *Threat Detection*.

Hypothesis 3b. *Threat Detection* mediates the effect Elaboration has on *Phishing Susceptibility*.

3.4. Attack Quality

Attack Quality is a new concept that is proposed in this study. It is used to describe various attack characteristics that are designed into a threat by an adversary in order to make it successful. This provides an additional dimension of study focused on the attacker that is currently lacking in research on phishing susceptibility as pointed out by Tetri & Vuorinen (2013) and Vishwanath et al. (2011). Attack Quality incorporates two constructs that are outlined in the Elaboration Likelihood Model; these are: Argument Quality and Persuasive Cues.

The Quality of Argument construct defines how well a position is justified based on available evidence or set of reasons. Using this criteria, a persuasive message is objectively judged based on its validity and merit. Luo et al. (2013) examine Quality of Argument in their study and show that users are likely to become victims if phishing messages have a high argument quality, possibly because they are less likely to detect the threat. This study examines the effect Quality of Argument has on both Threat Detection and Elaboration. It also examines the mediating effect that Threat Detection and Elaboration have on the overall Phishing Susceptibility. It is expected that phishing messages with high Quality of Argument will decrease Threat Detection and increase Elaboration. Therefore, this study hypothesizes that:

Hypothesis 4a. *Quality of Argument* has a negative effect on *Threat Detection*.

Hypothesis 4b. *Quality of Argument* has a positive effect on *Elaboration*.

Hypothesis 4c. *Threat Detection* mediates the effect *Quality of Argument* has on *Phishing Susceptibility*.

Hypothesis 4d. *Elaboration* mediates the effect *Quality of Argument* has on *Phishing Susceptibility*.

Persuasive cues are described as simple peripheral cues that are placed in a message in order to subjectively influence perceptions. Petty & Cacioppo (1986) propose that under low elaboration, people are influenced more by persuasive cues. This is because they do not actually expend effort processing the issue-relevant arguments. Various studies have examined the effect various persuasive cues have on susceptibility to phishing (Grazioli, 2004; Huber, Kowalski, Nohlberg, & Tjoa, 2009; Jakobsson, 2005; Karakasiliotis, Furnell, & Papadaki, 2006; Luo et al., 2013; Vishwanath et al., 2011; Wang et al., 2012; Workman, 2007, 2008a, 2008b). In these studies various persuasive cues have been enumerated; such as: spelling, grammar, layout, look and feel, security padlock icons, endorsements, logos, recipient-specific information, source, subject-line and genre-conformity. These cues have been found to have a significant effect in persuading people to trust deceptive messages. In many cases, these cues are an immediate way of communicating credibility without having to scrutinize the contents of a message. Attackers therefore design their attack with persuasive cues to defeat both Threat Detection and Elaboration. This study will also examine the mediating effect that Threat Detection and Elaboration have on Persuasive Cues and Phishing Susceptibility. Therefore, this study hypothesizes that:

Hypothesis 5a. *Persuasive Cues* have a negative effect on *Threat Detection*.

Hypothesis 5b. *Persuasive Cues* have a negative effect on *Elaboration*.

Hypothesis 5c. *Threat Detection* mediates the effect *Persuasive Cues* have on *Phishing Susceptibility*.

Hypothesis 5d. *Elaboration* mediates the effect *Persuasive Cues* have on *Phishing Susceptibility*.

3.5. Motivated to process

'Motivated to Process' is described by Petty & Cacioppo (1986) as the determination a person has to examine the content of a persuasive message. Two factors are thought to affect a person's motivation to process. These are: their level of involvement in the issue presented; and secondly their level of responsibility. Previous studies by Wang et al. (2012) and Vishwanath et al. (2011) have studied motivation to process only by examining involvement but not responsibility. This study, in addition, also examines responsibility as a factor. These previous studies have shown that the higher the Motivation to Process, the higher the Elaboration and objective scrutiny of the message for threats.

Involvement relates to personal relevance or vested interest on matters presented in the persuasive message. The more involved a person is, the more they will be motivated to process the message and also to detect threats. We will also examine the mediating effect that Threat Detection and Elaboration have on the relationship between Involvement and the overall Phishing Susceptibility. Therefore, this study hypothesizes that:

Hypothesis 6a. *Involvement* has a positive effect on *Threat Detection*.

Hypothesis 6b. *Involvement* has a positive effect on *Elaboration*.

Hypothesis 6c. *Threat Detection* mediates the effect *Involvement* has on *Phishing Susceptibility*.

Hypothesis 6d. *Elaboration* mediates the effect *Involvement* has on *Phishing Susceptibility*.

Responsibility refers to the obligation a person has to handle a matter and how accountable they are to its outcomes. The more accountable someone is to a matter, the more they will be motivated to process a message regarding it and also to detect threats. This study will also explore the mediating effect Threat Detection and Elaboration have on this relationship between Responsibility and the actual Phishing Susceptibility. Therefore, this study hypothesizes that:

Hypothesis 7a. *Responsibility* has a positive effect on *Threat Detection*.

Hypothesis 7b. *Responsibility* has a positive effect on *Elaboration*.

Hypothesis 7c. *Threat Detection* mediates the effect *Responsibility* has on *Phishing Susceptibility*.

Hypothesis 7d. *Elaboration* mediates the effect *Responsibility* has on *Phishing Susceptibility*.

3.6. Ability to process

The 'Ability to Process' is a concept that describes the capability an individual has to examine a persuasive message. An attacker's intention is to reduce an individual's ability to process using various techniques.

Petty & Cacioppo (1986) describe distraction as a factor that may affect a person's ability to process. In their work they explain that distractions require a person to exert more effort in order to examine a message. In fact, distractions often lead to low elaboration and a reliance on persuasive cues in making judgments. In addition, this study will examine the mediating effect that Threat Detection and Elaboration have on the relationship between Distractions and Phishing Susceptibility. Therefore, this study hypothesizes that:

Hypothesis 8a. *Distractions* have a negative effect on *Threat Detection*.

Hypothesis 8b. *Distractions* have a negative effect on *Elaboration*.

Hypothesis 8c. *Threat Detection* mediates the effect *Distractions* have on *Phishing Susceptibility*.

Hypothesis 8d. *Elaboration* mediates the effect *Distractions* have on *Phishing Susceptibility*.

Cialdini's (2001) six principles of influence and persuasion have been shown to have an impact on a person's ability to process social engineering threats. The six principles relate to authority, scarcity, liking and similarity, reciprocation, commitment and consistency and social proof. Studies by Karakasiliotis et al. (2006) and Workman (2007, 2008a, 2008b) have shown that these factors impair judgement and cause people to be more susceptible to social engineering attacks.

On careful examination, these six principles are seen to impair the ability to cognitively process in two ways: through emotions and pressure. The emotions that often come in play during social engineering attacks are fear, guilt and trust. Pressure is created by communicating a sense of urgency and by giving rewards or issuing ultimatums for a response to be given within a stipulated period of time. Luo et al. (2013) in their study hypothesized that time pressure reduces the ability to process. This study will also examine the mediating effect Threat Detection and Elaboration have on the relationship each of these have on Phishing Susceptibility. Therefore, this study hypothesizes that:

Hypothesis 9a. *Emotions* have a negative effect on *Threat Detection*.

Hypothesis 9b. *Emotions* have a negative effect on *Elaboration*.

Hypothesis 9c. *Threat Detection* mediates the effect *Emotions* have on *Phishing Susceptibility*.

Hypothesis 9d. *Elaboration* mediates the effect *Emotions* have on *Phishing Susceptibility*.

Likewise, with regards to Pressure this study hypothesizes that:

Hypothesis 10a. *Pressure* has a negative effect on *Threat Detection*.

Hypothesis 10b. *Pressure* has a negative effect on *Elaboration*.

Hypothesis 10c. *Threat Detection* mediates the effect *Pressure* has on *Phishing Susceptibility*.

Hypothesis 10d. *Elaboration* mediates the effect *Pressure* has on *Phishing Susceptibility*.

3.7. Knowledge

Knowledge is defined as the level of information and skills a person acquires (for example through experience or education) that affects their understanding of a matter. This concept has been studied widely in literature from various aspects. These aspects of knowledge include those relating to: terminology and threat techniques (the threat domain), cues that can be used to detect the threat (detection cues) and characteristics that can be used to distinguish legitimate communications (determinants of trust) (Dhamija et al., 2006; Downs et al., 2006; Fogg et al., 2001; Friedman, Hurley, Howe, Felten, & Nissenbaum, 2002; Garera, Provos, Chew, & Rubin, 2007; Grazioli, 2004; Jakobsson et al., 2007; Jakobsson & Ratkiewicz, 2006; Karakasiliotis et al., 2006; Sheng, Holbrook, Kumaraguru, Cranor, & Downs, 2010; Tsow & Jakobsson, 2007; Vishwanath et al., 2011; Wang et al., 2012).

Studies have shown that the more knowledge a person has regarding these various dimensions, the better their thought processing and also the more likely they are to correctly detect threats. This study examines the effect knowledge has on elaboration in line with the study by Wang et al. (2012) which showed that knowledge increased elaboration. This is because knowledge encouraged the processing of phishing indicators while weakening attention to deceptive triggers. However, unlike the study by Wang et al. (2012), this study sections the knowledge construct into its various dimensions with the intention of examining the effect each contributes. This study also examines the effect these knowledge dimensions have on Elaboration and Threat Detection and the overall Phishing Susceptibility. Therefore, this study hypothesizes that:

Hypothesis 11a. *Knowledge on Threat Domain* has a positive effect on *Threat Detection*.

Hypothesis 11b. *Knowledge on Threat Domain* has a positive effect on *Elaboration*.

Hypothesis 11c. *Threat Detection* mediates the effect *Knowledge on Threat Domain* has on *Phishing Susceptibility*.

Hypothesis 11d. *Elaboration* mediates the effect *Knowledge on Threat Domain* has on *Phishing Susceptibility*.

Likewise, with regards to Knowledge on Detection Cues this study hypothesizes that:

Hypothesis 12a. *Knowledge on Detection Cues* has a positive effect on *Threat Detection*.

Hypothesis 12b. *Knowledge on Detection Cues* has a positive effect on *Elaboration*.

Hypothesis 12c. *Threat Detection* mediates the effect *Knowledge on Detection Cues* has on *Phishing Susceptibility*.

Hypothesis 12d. *Elaboration* mediates the effect *Knowledge on Detection Cues* has on *Phishing Susceptibility*.

Likewise, with regards to Knowledge on Determinants of Trust this study hypothesizes that:

Hypothesis 13a. *Knowledge on Determinants of Trust* has a positive effect on *Threat Detection*.

Hypothesis 13b. *Knowledge on Determinants of Trust* has a positive effect on *Elaboration*.

Hypothesis 13c. *Threat Detection* mediates the effect *Knowledge on Determinants of Trust* has on *Phishing Susceptibility*.

Hypothesis 13d. *Elaboration* mediates the effect *Knowledge on Determinants of Trust* has on *Phishing Susceptibility*.

4. Methodology

The proposed model was analyzed with data collected from a naturalistic field study that involved staging phishing attacks at a university campus in Nairobi, Kenya. A key requirement for this research was to get an organization that would allow both the staging of a phishing attack and also the dissemination of the research results. Universities have been known to support similar research and were therefore a preferred choice for the study (Arachchilage & Love, 2013; Dodge et al., 2007; Finn & Jakobsson, 2007; Liang & Xue, 2010; Luo et al., 2013; Vishwanath et al., 2011; Wang et al., 2012; Workman, 2007, 2008a, 2008b).

A formal research application process was followed as outlined by the university's research office. The study was only carried out after the research proposal was reviewed and approved by the Institutional Review Board (IRB) and institutional consent was granted through the research office. In addition, technical oversight over the process was given through the university's Information Technology department. Two senior staff in the IT department were assigned to ensure that the phishing instruments did not cause any actual harm to the university or participants. The staged phishing attack did not collect or store any confidential information and neither did it transmit any malicious content in the process.

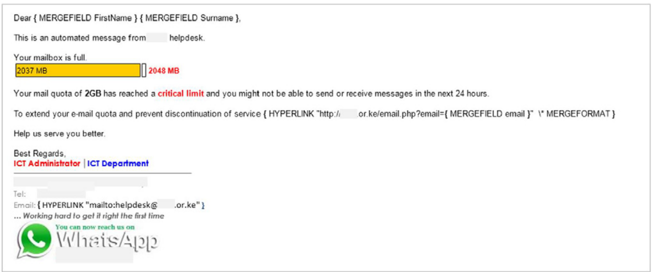
The study was allowed to proceed without alerting the university community about the research. This allowed the targeted insiders in the university to interact as they normally would with phishing attacks. It provided an assurance that the participants would not modify their behaviour by being aware of the ongoing study. This protected the study from the Hawthorne effect (Parsons, 1974). It also allowed the research to deliver a study with high ecological validity since it was set up to match real-world settings and everyday life for the population. This research design makes the results of this study generalizable to other similar contexts and populations (Huber et al., 2009; Workman, 2007).

It is recommended to use such a naturalistic field study methodology as opposed to the use of lab studies or phishing IQ tests. This is because lab studies and phishing IQ tests often do not match real-world settings and can be plagued by bias introduced by the Hawthorne effect

(Anandpara, Dingman, Jakobsson, Liu, & Roinestad, 2007; Dhamija et al., 2006). Lab studies usually heighten the awareness of the participants and this may cause them to change their behaviour contrary to what it would have been in a natural setting. In addition, these studies have required participants to volunteer for the research. There could be unique characteristics about those who volunteer for the study that distinguishes them from the wider population (Kumaraguru et al., 2009). These reasons could make the results of such studies harder to generalize.

The development of the phishing instruments for the study was guided by the recommendations and lessons learnt from previous studies by Luo et al. (2013), Arachchilage and Love (2013), Vishwanath et al. (2011) and Bakhshi, Papadaki, and Furnell (2009). Firstly, samples of actual phishing attacks that had been recently targeted at the university were studied. This allowed various characteristics that made the phishing attacks convincing to be identified with respect to previous work by Dhamija et al. (2006), Downs et al. (2006), Downs et al. (2007), Jakobsson et al. (2007) and Kumaraguru, Sheng, et al. (2007). Secondly, care was taken to identify the structure, look and feel of normal communication at the university so as to provide genre conformity. This was to ensure that the phishing would not be dismissed immediately. It was to encourage the targeted insiders to go beyond the first glance and allow for actual interaction and thought processing. Thirdly, a topic of interest was selected to ensure that the staged phishing content would be captivating and of interest to the targeted insiders. This form of pretexting is also required to ensure targeted insiders process the phishing communication. It was noted that users often received email alerts informing them that their mailbox was full. This message line was therefore chosen and a sense of urgency was also designed into the phishing attack by requiring the targeted user to respond quickly so as to prevent discontinuation of email services.

Next was the development of the phishing instruments. A domain that imitates the university's domain was purchased. The university's domain ended with ac.ke and the phishing domain that was registered ended with or.ke. The purchase of the domain also allowed an email address helpdesk@uni.or.ke to be set up to imitate the legitimate email address used by the IT team to communicate to the users. In addition, the name descriptor associated with this email address was set to be identical to the legitimate one thereby spoofing it. This allowed the phishing emails look as legitimate as possible. Next the phishing email was authored and setup using mail merge on Microsoft Word 2013 as shown in Fig. 3. The mail merge template allowed the phishing emails to be customized using first name, surname and for the targeted user email address to be passed on as a GET parameter to the phishing website. This made the phishing to be personalized as would be in a spear phishing campaign. The phishing email matched the look and feel of communication sent by the university and also was professionally authored to avoid sloppy content or grammatical mistakes. This was to prevent it from being dismissed at first-glance and to encourage users to take it seriously enough to process it. Phishing emails were automatically sent by running the mail merge against the list of targeted insiders and integrating this process with Microsoft Outlook.



Dear { MERGEFIELD FirstName } { MERGEFIELD Surname },

This is an automated message from helpdesk.

Your mailbox is full.

Your mail quota of 2GB has reached a critical limit and you might not be able to send or receive messages in the next 24 hours.

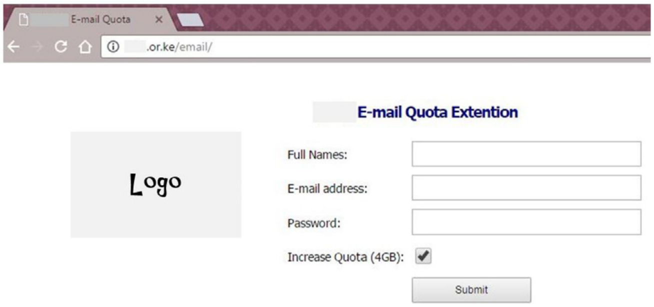
To extend your e-mail quota and prevent discontinuation of service { HYPERLINK "http:// .or.ke/email.php?email={ MERGEFIELD email }" * MERGEFORMAT }

Help us serve you better.

Best Regards,
ICT Administrator | ICT Department

Tel:
Email: { HYPERLINK "mailto:helpdesk@ .or.ke" }
... *Working hard to get it right the first time*

Fig. 3. Phishing Mail Merge Template.



E-mail Quota Extention

Full Names:

E-mail address:

Password:

Increase Quota (4GB): ☑

Fig. 4. Phishing Webpage.

A phishing site was also developed using HTML5, CSS and PHP with a backend MySQL database. It was hosted on the or.ke registered domain and tested to ensure it ran without errors. The IT staff assigned to the study reviewed all code and carefully examined the data used to ensure that no confidential data was captured, transmitted or stored. Fig. 4 shows the login page that was developed for the phishing exercise.

These phishing instruments collected various data items that measured phishing susceptibility.

Delivery notifications in Microsoft Outlook tracked when the email was successfully delivered and also when the email was opened. This process could have been automated by integrating hidden scripts in the email but the use of delivery notifications afforded end users transparency in the process. The phishing email also had a hyperlink with the words “click here” highlighted in blue and underlined. When clicked, this hyperlink did two things. First, it redirected the person to the phishing page by loading the phishing address on their default browser. Secondly, it passed on the email address as a pre-filled parameter to populate the phishing webpage. This made it possible to uniquely track each person who loaded the phishing page.

The phishing page also ran background scripts that recorded the timestamp, user identifier and various system parameters including the operating system, browser model and IP address. This allowed data to be automatically collected even when the user did not interact with the phishing page. Loading the phishing page was sufficient to indicate phishing susceptibility and provide valuable information.

Phishing susceptibility was also measured by observing interactions with the phishing page. If the page was loaded from the hyperlink on the phishing email, the person's email address was pre-filled. The phishing page also requested the person's full names and password. If a person clicked the submit button the password was neither captured nor transmitted in order to protect the institution and individuals from actual harm. Error validation also ensured that the page would not be submitted if the fields were blank.

The probability sampling technique was used to select a sample of 4483 insiders from the population of 8405. Table 3 outlines the sampling frame and the resulting sample. The actual insiders were then selected from the list using simple random sampling and their emails were loaded on mail merge list.

The staged phishing attack ran for 40 days. Within this time all the

Table 3
Population Sampling.

Strata	Sampling Frame	Proportion	Size in Sample
Students	7729	91.96%	4122
Staff	312	3.71%	166
Adjunct Faculty	158	1.88%	84
Full-time Faculty	141	1.68%	75
Management	13	0.15%	6
Interns	9	0.11%	4
Mailing List Users	7	0.08%	7
Unknown	36	0.43%	19
Total	8405	100%	4483

Table 4
Operationalization of Study Variables.

Variables	Measurement Items	Informing Literature	Cronbach's Alpha
Questionnaire Self-Reported on Phishing Susceptibility (Dependent Variable) Directly Observed Behaviour on Phishing Susceptibility (Dependent Variable)	QSR_DV1: Did you read this email? QSR_DV2: Did you click the link labelled "click here" on this email? QSR_DV3: Did you fill in the form presented on the website? DOB_DV1: Observed click behaviour from the website DOB_DV2: Observed form-fill behaviour from the website	- Liang and Xue (2010) - Workman (2007) - Bakhshi et al. (2009)	0.823
Threat Detection	TD1: I could tell this was an online attack TD2: I could tell someone was trying to deceive me TD3: I could tell that someone was trying to capture my personal details and password	- Arachchilage and Love (2013) - Liang and Xue (2010)	0.926
Elaboration	ELAB1: I made conscious effort to evaluate the email/website ELAB2: I took time to evaluate the email/website ELAB3: I carefully evaluated the email/website	- Wang et al. (2012) - Vishwanath et al. (2011) - Petty & Cacioppo (1986)	0.937
Argument Quality	QA1: I carefully scrutinized the email message before responding QA2: I reasoned through the explanation given in the email before responding QA3: I examined the reasons given in the email before responding	- Luo et al. (2013) - Petty & Cacioppo (1986)	0.896
Persuasive Cues	Please rate to which extent the following components of the email/website influenced your response PC1: Source credibility (i.e. ICT administrator) PC2: Personalized Greeting PC3: Offer to extend your mail quota PC4: Warning that your email service would be discontinued PC5: Urgency to respond within 24 h PC6: Resemblance to other [UNI] emails PC7: Resemblance to other [UNI] websites	- Luo et al. (2013) - Wang et al. (2012) - Vishwanath et al. (2011) - Workman (2007)	0.898
Involvement	INV1: The email seemed very relevant to me INV2: The email seemed very important to my work/studies INV3: The email seemed very applicable to my current situation	- Wang et al. (2012) - Vishwanath et al. (2011) - Petty & Cacioppo (1986)	0.925
Responsibility	RES1: I am answerable to communications I receive on my [UNI] email account RES2: I am in control of the day-to-day operation of my [UNI] email account RES3: I consider myself responsible for my [UNI] email account	- Petty & Cacioppo (1986)	0.790
Distraction	DIST1: There is usually a lot of activity going on around me when reading and responding to emails DIST2: I usually multi-task when reading and responding to emails DIST3: I tend to be distracted when reading and responding to emails	- Petty & Cacioppo (1986)	0.761
Emotions	EM1: Reading the email invoked an emotion in me (e.g. fear, anxiety) EM2: I responded to this email so that I would not get into trouble EM3: I would have felt guilty for not responding to the email	- Workman (2007)	0.763
Pressure	PRES1: I am usually under pressure to move on to other tasks when reading and responding to emails PRES2: I usually have a sense of urgency when reading and responding to emails PRES3: I tend to rush through my emails	- Luo et al. (2013) - Vishwanath et al. (2011)	0.753
Knowledge Quiz	KQC: Knowledge Quiz Count value is an integer sum of correct answers to the six questions below and 0 ≤ KQC ≤ 6 Questions: Please indicate what the following words mean with regards to information security KQ1: Phishing KQ2: Social Engineering KQ3: URL KQ4: Certificate KQ5: Spoofing KQ6: Domain *Options to select from:* A: I have never seen this word before B: I have seen this word before but I don't know what it means C: A file used to identify websites and encrypt data D: Manipulating people to compromise the security of their systems E: A name that identifies an organization's resources on the internet F: Forging the identity of a trusted entity G: Impersonation commonly through email that tricks people into sharing sensitive information H: A term for insecure websites I: Malicious Software J: A web address	- Vishwanath et al. (2011) - Downs et al. (2006)	0.744
Knowledge on Threat Domain	KTD1: I have sufficient knowledge regarding this type of threat KTD2: I have sufficient knowledge regarding the consequences of this type of threat KTD3: I have sufficient knowledge on how to detect this type of	- Vishwanath et al. (2011) - Downs et al. (2006) - Downs et al. (2006)	0.744

(*continued on next page*)

Table 4 (*continued*)

Variables	Measurement Items	Informing Literature	Cronbach's Alpha
	threat KTD4: I have sufficient knowledge on how to respond to this type of threat		
Knowledge on Detection Cues	KDC1: I know how to reveal hyperlinks hidden behind text to detect such threats KDC2: I know how to analyze web addresses to detect such threats KDC3: I know how to analyze web certificates to detect such threats	- Garera et al. (2007) - Downs et al. (2007) - Downs et al. (2006) - Dhamija et al. (2006)	0.938
Knowledge on Determinants of Trust	To what extent did you use the following characteristics or techniques to determine the trustworthiness of the email/website? KDT1: Consistency in logo, colors, look and feel KDT2: Grammar and Spelling KDT3: Personalized greeting with your names KDT4: Content (e.g. reasonableness of the explanation in email and website content) KDT5: Context (e.g. it was expected in the prevailing circumstances) KDT6: Email address of the sender KDT7: Contacting the [UNI] ICT helpdesk KDT8: Asking someone (e.g. colleague, friend) KDT9: Web address and hyperlink evaluation KDT10: Website encryption or padlock icon KDT11: Website certificate KDT12: Domain registration information (e.g. from whois) KDT13: Security tool information (e.g. anti-phishing tool integrated in email/browser)	- Tsow & Jakobsson (2007) - Dhamija et al. (2006) - Karakasiliotis et al. (2006)	0.882
CONTROL VARIABLES			
Gender	GENDER: What is your gender?	Sheng et al. (2010)	N/A
Age	AGE: What is your age in years?	Bulgurcu, Cavusoglu, and Benbasat (2010)	N/A
Level of Education	EDUCATION: What is the highest level of education you have completed?	Sheng et al. (2010)	N/A
Role	ROLE: What is your role at the university?	Kumaraguru et al. (2009)	N/A
Year first used the internet	YEAR_INTERNET: Which year did you first use the internet?	Sheng et al. (2010)	N/A
Hours spent on the internet in a day	HOURS_INTERNET: How many hours do you spend on the internet in a day?	Kumaraguru, Sheng, et al. (2007)	N/A
Computer Skills	COMP_SKILLS: How would you rate your computer skills?	Bulgurcu et al. (2010)	N/A
Email Load	EL: How many emails do you receive in your official email account in a day?	Vishwanath et al. (2011)	N/A
Email Responsiveness	ER1: I *read* all emails I receive in my [UNI] official email account ER2: I *respond* to all emails I need to in my [UNI] official email account	Vishwanath et al. (2011)	N/A
Online Services	To what extent do you use the following online services? OS1: Email OS2: Social Media OS3: Online Shopping OS4: Online Banking	- Downs et al. (2007) - Downs et al. (2006)	N/A
Prior Victimization	Have you ever experienced the following online threats in the past? PV1: Scam PV2: Online Account Hijacking PV3: Identity Theft PV4: Credit/Debit Card Fraud PV5: Malicious software infection	- Workman (2008b) - Downs et al. (2006)	N/A
Risk Propensity	To what extent do you agree with the following statements about your risk propensity? RP1: I like taking risks RP2: People say I am a risk taker RP3: I sometimes take risks that could threaten my safety	- Sheng et al. (2010) - Downs et al. (2006)	0.762

4483 insiders were sent the phishing email once and then data collected based on their interactions with the phishing email and webpage. The exercise was stopped when a student, who is a prominent social blogger, posted an alert on social media calling for all users to be on alert and for the university to investigate the attack. This prompted the university to call off the exercise and the IT director sent out an email informing the community of the study aimed at addressing phishing susceptibility.

A questionnaire was used to measure the study's antecedent variables because, unlike the outcome variable, they could not be directly observed. When handing out questionnaires, the respondents were not informed whether they were victims of the phishing exercise or not. Study participants had to self-report their internal perceptions and thought processes. The questionnaire measurement items and the reference studies used to develop them are as outlined in Table 4.

5. Results

Based on the stratified probability sampling of the university population, 4483 phishing emails were sent during this study. Delivery failures were received for 138 email addresses. No read receipts or sign of interaction was noted from 4104 email addresses. This was a majority of those targeted in the study. Only 241 insiders were observed to have opened the phishing email. These 241 insiders are hereafter termed as the "active participants" because they actively engaged with the phishing stimulus chosen as the basis of assessing phishing susceptibility this study. The active participants are a 5.37% representation of those targeted. This percentage may be considered to be low but is comparable to that obtained in a similar study by Mohebzada, El Zarka, BHojani, & Darwish (2012) where two types of phishing emails were sent to a university population. Over 10,000 phishing emails were sent and the first phishing email had an 8.74% success rate while the second had only 2.05% success rate.

Data from the backend phishing database showed a total of 98 clicks on the phishing email hyperlink associated with 75 unique users since some clicked the phishing hyperlink multiple times. In addition, the phishing webpage was filled in 72 times with 66 of these form-fills being unique and 6 being repeated entries. This means that 31.12% of the participants were susceptible to phishing. Of these, 88% disclosed passwords that would enable an attacker gain access to the organization's systems.

Data was also collected from the 241 who participated in the exercise through questionnaires. A total of 192 filled-in questionnaires were collected and their data used to test the proposed model. This represents a 79.67% response rate from administered questionnaires.

The data was screened for missing values, outliers, common method bias and assumptions of normality. Twenty-one questionnaires had some missing data.

The missing data per questionnaire ranged from 0.7% to 8.2%. The missing data per variable ranged from 0.5% to 1.6%. These were less than the thresholds of 10% per questionnaire and 15% per variable set by Hair, Black, Babin, and Anderson (2009). Therefore, imputation for missing values was done and replacement values were calculated.

The Mahalanobis D^2 measure was computed and used to detect outliers for this study. No outliers were detected since none of the p-values of the Mahalanobis Chi-Square Cumulative Distribution Function was less than 0.001.

Common Method Bias was assessed using the Harman one-factor test. An un-rotated factor analysis was done to extract only one factor. The resulting factor was found to explain only 17.13% of the total variance. This was less than the threshold of 50% set by Podsakoff, MacKenzie, Lee, and Podsakoff (2003). Therefore, Common Method Bias was not considered a threat for the data set.

Normality was analyzed using the skewness and kurtosis values for each measure. The guideline for test of normality from Curran, West, and Finch (1996) was to identify measures whose skewness values are greater than 2 and kurtosis values are greater than 3. None of the variables had normality issues using this criteria.

An analysis of the demographic distribution showed that the majority of the respondents were male (63%), members of staff at the university (38.5%), in the 26–35 year age bracket (34.4%) and had completed a graduate degree (36.5%) as outlined in Table 5.

5.1. Exploratory Factor Analysis (EFA)

IBM SPSS Statistics version 23 was used to perform the Exploratory Factor Analysis using the dataset of 192 cases. The KMO and Bartlett's test of sphericity and the reproduced correlation matrix were derived in the factor analysis. Factor extraction was done using the Maximum Likelihood method because it maximizes the difference between factors while also providing many helpful indices for assessing the resulting factor structure; particularly indices that can evaluate the goodness-of-fit. No predefined number of factors was defined for extraction; the extraction was based on Eigenvalues greater than 1. Promax rotation method was selected because it is the widely used orthogonal rotation method (Kline, 2013). The Kappa value was set to 4. In addition, small loadings less than ± 0.3 were suppressed with the objective of obtaining a clean pattern matrix. According to Hair et al. (2009) factor loadings of ± 0.3 to ± 0.4 are minimally acceptable although values greater than ± 0.5 are more desirable. An iterative process was followed to remove cross loadings and measures with small loadings. The measures removed were EM1, EM3, EM2 and DT7 respectively. This meant that the Emotions variable was completely dropped when cleaning the pattern matrix.

The resulting pattern matrix is as shown in Table 6. Factor 1 combined the Knowledge on Threat Domain and Knowledge on Detection Cues into one factor (KTDC). Knowledge on Determinants of Trust was split into two factors and after analysis these were found to represent Low Determinants of Trust (factor 4) and High Determinants of Trust (factor 3). Factor 6 combined Distraction and Pressure variables into one factor. This meant that the Ability to Process (AP) factors combined into one dimension.

A total of 11 factors were extracted in the Exploratory Factor Analysis. The Kaiser-Meyer-Olkin Measure of Sampling Adequacy was 0.854 (above 0.7 threshold), the total variance explained was 67.471% (above 60% threshold), and the Goodness of Fit using Chi-Square/df was 1.72 (within the range of 1 and 3). All these measures indicated that the resulting factor matrix was good.

5.2. Confirmatory Factor Analysis (CFA)

Next, the Confirmatory Factor Analysis was done on IBM AMOS version 23 using the dataset of 192 cases. The measurement model was drawn from the factors extracted from the EFA. Measurement loadings were above 0.5 indicating good unidimensionality. Reliability was analyzed using the Composite Reliability (CR) score. These values were above the 0.7 threshold indicating good measurement reliability (Nunnally & Bernstein, 1994). Convergent validity was assessed using the Average Variance Extracted (AVE) measure with a desired threshold of 0.5 or more for each factor (Fornell & Larcker, 1981; Hair et al., 2009). The Ability to Process (AP) factor did not have adequate AVE. This necessitated DIST1, DIST2 and DIST3 measures to be dropped, one at a time, in order to meet the required threshold. This effectively meant that the Distraction factor was removed from the model and only the Pressure factor remained in the Ability to Process construct. Discriminant validity was assessed using the square root of AVE. The square root of AVE for each factor was found to be greater than any correlation between factors in the model indicating good discriminant validity (Fornell & Larcker, 1981). Table 7 shows the results of the reliability and validity analysis. There were no reliability or validity concerns.

The resulting CFA measurement model fit was found to be satisfactory using Goodness-of-Fit Measures outlined in Table 8. Desired thresholds are specified by Hu and Bentler (1999).

5.3. Structural model

After a satisfactory measurement model was obtained, the SEM path

Table 5
Demographics.

Gender	Role at the University	Age in Years	Highest level of Education
Male (63%)	Student (38.0%)	18–25 years (23.4%)	Primary School (0%)
Female (37%)	Faculty/Lecturer (23.4%)	26–35 years (34.4%)	High School (13.5%)
	Staff (38.5%)	36–45 years (22.9%)	Diploma (5.7%)
		46–55 years (10.4%)	Undergraduate Degree (30.7%)
		Above 55 years (8.9%)	Graduate Degree (36.5%)
			Doctoral Degree (13.5%)

Table 6
Exploratory Factor Analysis Pattern Matrix.

	Factor[a]										
	1 (KTDC)	2 (PC)	3 (KDT_HIGH)	4 (KDT_LOW)	5 (DV)	6 (AP)	7 (INV)	8 (ELAB)	9 (QA)	10 (RES)	11 (TD)
KTD1	.918										
KTD3	.889										
KTD4	.874										
KTD2	.839										
KDC2	.823										
KDC1	.789										
KDC3	.739										
PC5		.997									
PC4		.977									
PC7		.658									
PC3		.623									
PC2		.576									
PC6		.565									
PC1		.502									
KDT11			.937								
KDT13			.840								
KDT12			.800								
KDT10			.798								
KDT9			.682								
KDT8			.433								
KDT4				.838							
KDT1				.724							
KDT2				.706							
KDT5				.683							
KDT3				.682							
KDT6				.596							
DOB_DV2					1.004						
DOB_DV3					.955						
QSR_DV2					.528						
QSR_DV3					.392						
DIST3						.707					
DIST1						.696					
PRES1						.690					
DIST2						.666					
PRES3						.640					
PRES2						.631					
INV2							.917				
INV1							.908				
INV3							.877				
ELAB2								.916			
ELAB1								.867			
ELAB3								.782			
QA3									.937		
QA2									.856		
QA1									.666		
RES3										.840	
RES2										.796	
RES1										.666	
TD2											.862
TD3											.794
TD1											.733

Extraction Method: Maximum Likelihood.
Rotation Method: Promax with Kaiser Normalization.

[a] Rotation converged in 8 iterations.

model based on the proposed conceptual model was then constructed in IBM AMOS version 23. This was with the modifications done during the EFA and CFA. The Emotions and Distractions factors were removed; the Knowledge of Threat Domain and Detection Cues factors were combined into one, and the Knowledge on Determinants of Trust was split into two factors representing High and Low Determinants of Trust.

The model was able to account for 69.5% of the variance in Threat Detection, 50.8% of variance in Elaboration and 28% of the variance in Phishing Susceptibility.

The final model after the Structural Equation Model analysis is illustrated in Fig. 5.

The structural model demonstrated decent Goodness-of-Fit as shown in measures outlined in Table 9.

5.4. Control variables

Various control variables were added to the model and their effects on Phishing Susceptibility were analyzed. A total of 12 controls were examined which were: Gender, Age, Level of Education, Role, Years on Internet, Hours on Internet, Computer Skill, Email Load, Email Responsiveness, Online Service Usage, Prior Victimization and Risk Propensity. Only three controls were found to have a significant effect on the dependent variable and were therefore retained in the final model. These were: Role at the University, Email Load and Email Responsiveness.

Results showed that staff and faculty were more susceptible to phishing than students. Contrary to expectations regarding email load put forth by Vishwanath et al. (2011), individuals who received high

Table 7
Reliability, Convergent Validity and Discriminant Validity Analysis.

	CR	AVE	MSV	MaxR(H)	DV	TD	ELAB	QA	PC	INV	RES	PRES	KTDC	KDT_LOW	KDT_HIGH
DV	0.868	0.633	0.23	0.944	**0.796**										
TD	0.928	0.811	0.534	0.945	−0.479***	**0.901**									
ELAB	0.939	0.837	0.405	0.95	−0.226**	0.452***	**0.915**								
QA	0.898	0.747	0.405	0.903	−0.094	0.362***	0.636***	**0.864**							
PC	0.894	0.56	0.317	0.948	0.342***	−0.460***	−0.017	0.093	**0.749**						
INV	0.926	0.807	0.298	0.931	0.413***	−0.496***	−0.054	0.06	0.546***	**0.898**					
RES	0.81	0.591	0.075	0.846	−0.092	0.150†	0.273**	0.268**	0.017	0.072	**0.769**				
PRES	0.758	0.514	0.021	0.787	−0.05	−0.034	−0.062	−0.144†	0.071	0.056	−0.018	**0.717**			
KTDC	0.942	0.699	0.534	0.962	−0.362***	0.730***	0.422***	0.389***	−0.377***	−0.371***	0.214*	−0.048	**0.836**		
KDT_LOW	0.873	0.537	0.317	0.888	0.160*	−0.113	0.176*	0.170*	0.563***	0.325***	0.108	0.011	−0.035	**0.733**	
KDT_HIGH	0.899	0.603	0.232	0.934	−0.277***	0.450***	0.389***	0.416***	0.009	−0.249**	0.151†	−0.057	0.482***	0.363***	**0.777**

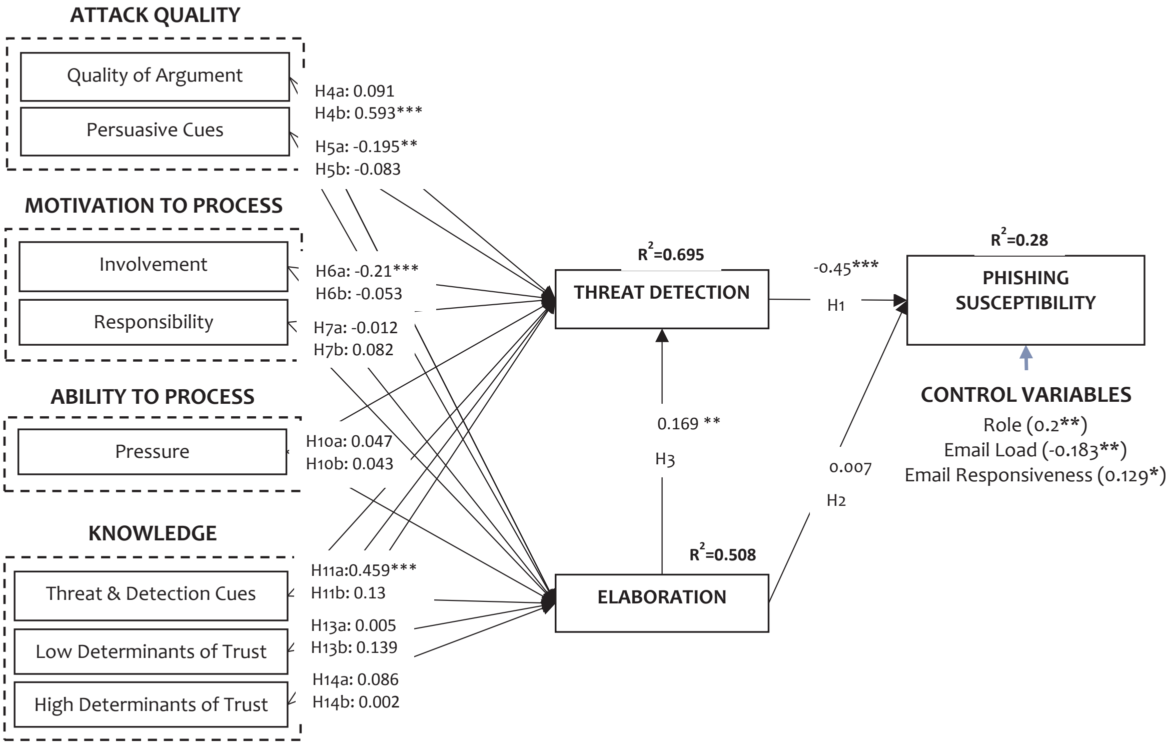
NOTES:
CR stands for Composite Reliability. CR should be greater than 0.7 and should also be greater than the AVE (Nunnally & Bernstein, 1994).
AVE stands for Average Variance Extracted. AVE should be greater than 0.5 (Fornell & Larcker, 1981; Hair et al., 2009).
MSV stands for Maximum Shared Variance. MSV should be less than Average Variance Extracted (AVE) (Hair et al., 2009).
MaxR(H) stands for Maximum Reliability.
After MaxR(H) column is the correlation matrix with the square root of AVE on the diagonal (values in bold).
Square root of AVE values (in bold on the diagonal) should be larger than off-diagonal correlation values (Fornell & Larcker, 1981).
Significance of Correlations: *** $p < 0.001$; ** $p < 0.010$; * $p < 0.050$; † $p < 0.100$.

Table 8
Measurement Model Goodness-of-Fit.

Goodness-of-Fit Measures	Notation	Desired Threshold	Value Obtained	Remark
Chi-square	χ^2		1461.465	
Degrees of freedom	df		1010	
Chi-square/df ratio	χ^2/df	1 to 3	1.447	Excellent
Comparative Fit Index	CFI	> 0.90	0.938	Good
Standardized Root Mean Square Residual	SRMR	< 0.08	0.067	Excellent
Root Mean Square Error of Approximation	RMSEA	< 0.05	0.048	Excellent
p of Close Fit	PCLOSE	> 0.05	0.684	Excellent

Table 9
Structural Model Goodness-of-Fit.

Goodness-of-Fit Measures	Notation	Desired Threshold	Value Obtained	Remark
Chi-square	χ^2		59.517	
Degrees of freedom	df		37	
Chi-square/df ratio	χ^2/df	1 to 3	1.609	Excellent
Comparative Fit Index	CFI	> 0.90	0.976	Good
Standardized Root Mean Square Residual	SRMR	< 0.08	0.07	Excellent
Root Mean Square Error of Approximation	RMSEA	< 0.05	0.056	Excellent
p of Close Fit	PCLOSE	> 0.05	0.323	Excellent



Notes: *** p-value < 0.001; ** p-value < 0.010; * p-value < 0.050; ns = not significant

Fig. 5. Structural Equation Model Analysis.

volumes of email were less susceptible to phishing. It was expected that increased email load would lead to habitual responses and lack of attention which in turn would lead to increased susceptibility. However, this was not the case. It could be those who received many emails did not process many of them. Therefore it was important to examine how responsive these individuals were to their emails. Results showed that the individuals who were most responsive to their emails were the ones who were most susceptible to phishing attacks.

5.5. Hypothesis testing

The first round of Hypothesis testing did not examine moderation or mediation effects. In this first round, a total of 25 relationships were tested of which: 6 were found to be significant, 6 were dropped at CFA or EFA level and 13 were not significant. Hypothesis 1, 3a, 4b, 5a, 6a and 11a were supported because their standardized effects were found to be significant. However, Hypothesis 2, 4a, 5b, 6b, 7a, 7b, 10a, 10b, 11b, 13a, 13b, 13c and 13d were not supported because their standardized effects were not found to be significant at the $p \le 0.05$ level. Hypothesis 8a and 8b were removed because the Distraction variable was dropped during CFA. Hypothesis 9a and 9b were also removed because Emotions variable was dropped during EFA. Variables in Hypothesis 11 and 12 were merged to become Knowledge of Threat and Detection Cues (KTDC) since they were merged into one dimension during EFA. Therefore, Hypothesis 11a and 11b were used to represent both and Hypothesis 12a and 12b were removed.

The results of the Hypothesis testing are summarized in Table 10.

Hypothesis 1 posited that *Threat Detection has a negative effect on Phishing Susceptibility*. As hypothesized, results showed a significant negative effect ($\beta = -0.45$, $p < 0.001$). Hence the data supported Hypothesis 1. This means that as threat detection increases, phishing susceptibility decreases.

Hypothesis 2 posited that *Elaboration has a negative effect on Phishing Susceptibility*. Results showed a negative effect (in line with the hypothesized direction) but this effect was not significant ($\beta = -0.007$, $p = 0.922$). This means that increased elaboration does lead to decreased phishing susceptibility but this effect was too small and not supported, beyond chance, by the data.

Hypothesis 3a posited that *Elaboration has a positive effect on Threat Detection*. Results showed a significant positive effect ($\beta = 0.169$, $p = 0.003$). This means that as elaboration increases, threat detection also increases. Thus, hypothesis 3a was supported by the data.

Hypothesis 4a posited that *Quality of Argument has a negative effect on Threat Detection*. However, results showed a positive effect (contrary to hypothesized direction) but this effect was not significant ($\beta = 0.091$, $p = 0.136$). It was initially hypothesized that attackers would craft messages with high argument quality so as to evade threat detection. However results showed a different effect; linking increased argument quality with increased threat detection. Hypothesis 4a was not supported by the data.

Hypothesis 4b posited that *Quality of Argument has a positive effect on Elaboration*. Results showed a significant positive effect ($\beta = 0.593$, $p < 0.001$). This means that as the quality of argument increases, elaboration also increases. This effect was the strongest among all the tested antecedent relationships that lead to elaboration. Hypothesis 4b was supported by the data.

Hypothesis 5a posited that *Persuasive Cues have a negative effect on Threat Detection*. Results showed a significant negative effect ($\beta = -0.195$, $p = 0.002$). This means that an increase in persuasive cues leads to a decrease in threat detection. Hypothesis 5a was supported by the data.

Hypothesis 5b posited that *Persuasive Cues have a negative effect on Elaboration*. Results showed a negative effect (in line with the hypothesized direction), but this effect was not significant ($\beta = -0.083$, $p = 0.289$). This means that an increase in persuasive cues showed a decrease in elaboration but this effect was not strong enough to support the hypothesis beyond chance. Hypothesis 5b was not supported by the data.

Hypothesis 6a posited that *Involvement has a positive effect on Threat Detection*. Results showed a significant negative effect (contrary to hypothesized direction) ($\beta = -0.21$, $p < 0.001$). The initially hypothesized stance was that the more a person was involved on a subject matter touched on by phishing communication, the more likely they would be to detect the threat. However this was not the case. Results showed that the more a person is involved on a subject matter communicated in a phishing message, the less likely they are to detect a

Table 10
Hypothesis Testing Results.

Hypothesis	Path	Comment	β values	S.E.	C.R.	p-value	Conclusion
H1	TD - > DV		−0.45	0.022	−6.463	***	Supported
H2	ELAB - > DV		−0.007	0.023	−0.097	0.922	Not Supported
H3a	ELAB - > TD		0.169^{**}	0.06	2.965	0.003	Supported
H4a	QA - > TD		0.091	0.068	1.49	0.136	Not Supported
H4b	QA - > ELAB		0.593^{***}	0.069	9.198	***	Supported
H5a	PC - > TD		-0.195^{**}	0.059	−3.175	0.002	Supported
H5b	PC - > ELAB		−0.083	0.071	−1.061	0.289	Not Supported
H6a	INV - > TD		-0.21^{***}	0.058	−3.956	***	Supported
H6b	INV - > ELAB		−0.053	0.069	−0.784	0.433	Not Supported
H7a	RES - > TD		−0.012	0.064	−0.274	0.784	Not Supported
H7b	RES - > ELAB		0.082	0.077	1.534	0.125	Not Supported
H8a	DIST - > TD	Dropped at CFA	–	–	–	–	Removed
H8b	DIST - > ELAB	Dropped at CFA	–	–	–	–	Removed
H9a	EM - > TD	Dropped at EFA	–	–	–	–	Removed
H9b	EM - > ELAB	Dropped at EFA	–	–	–	–	Removed
H10a	PRES - > TD		0.047	0.061	1.139	0.255	Not Supported
H10b	PRES - > ELAB		0.043	0.073	0.824	0.41	Not Supported
H11a	KTDC - > TD	Merged KTD with KDC at EFA	0.459^{***}	0.064	8.052	***	Supported
H11b	KTDC - > ELAB	Merged KTD with KDC at EFA	0.13	0.076	1.818	0.069	Not Supported
H12a	KDC - > TD	Merged with KTD at EFA	–	–	–	–	Removed
H12b	KDC - > ELAB	Merged with KTD at EFA	–	–	–	–	Removed
H13a	KDT_LOW - > TD	Split from KDT at EFA	0.005	0.077	0.087	0.93	Not Supported
H13b	KDT_LOW - > ELAB	Split from KDT at EFA	0.139	0.092	1.851	0.064	Not Supported
H14a	KDT_HIGH - > TD	Split from KDT at EFA	0.086	0.061	1.478	0.139	Not Supported
H14b	KDT_HIGH - > ELAB	Split from KDT at EFA	0.002	0.074	0.028	0.977	Not Supported

Notes: *** p-value < 0.001; ** p-value < 0.010; * p-value < 0.050.

phishing threat. Hypothesis 6a was supported by the data but for an inverse relationship.

Hypothesis 6b posited that *Involvement has a positive effect on Elaboration*. However, results showed a negative effect (contrary to hypothesized direction) but this effect was not significant (β = −0.053, p = 0.433). The earlier hypothesized stance was that the more someone was involved on an issue communicated in a phishing message, the keener they would be to cognitively process the message. However, results showed that the more the involvement, the lower the cognitive processing. Nonetheless, this effect was small and could not be argued beyond chance. Therefore, hypothesis 6b was not supported by the data.

Hypothesis 7a posited that *Responsibility has a positive effect on Threat Detection*. However, results showed a negative effect (contrary to hypothesized direction) but this effect was not significant (β = −0.012, p = 0.784). The earlier hypothesized position was that the more responsible someone was regarding a subject matter, the more likely they were to detect a phishing threat. However, results showed different. This means that the higher a person's sense of responsibility on a subject matter, the less likely they were to detect a phishing threat. However, this effect was too small to be attributed beyond chance. Hypothesis 7a was not supported by the data.

Hypothesis 7b posited that *Responsibility has a positive effect on Elaboration*. Results showed a positive effect (in line with the hypothesized direction) but this effect was not significant (β = 0.082, p = 0.125). This means that the more responsible someone was regarding a subject matter, the more likely they were to cognitively process the message. However, this effect was not strong enough to support the hypothesis beyond chance. Therefore, hypothesis 7b was not supported by the data.

Hypothesis 8a posited that *Distractions have a negative effect on Threat Detection*. This hypothesis was not tested because the Distractions construct was dropped during Confirmatory Factor Analysis.

Hypothesis 8b posited that *Distractions* have a negative effect on *Elaboration*. This hypothesis was not tested because the Distractions construct was dropped during Confirmatory Factor Analysis.

Hypothesis 9a posited that *Emotions* have a negative effect on *Threat Detection*. This hypothesis was not tested because the Emotions construct was dropped during Exploratory Factor Analysis.

Hypothesis 9b posited that *Emotions* have a negative effect on *Elaboration*. This hypothesis was not tested because the Emotions construct was dropped during Exploratory Factor Analysis.

Hypothesis 10a posited that *Pressure has a negative effect on Threat Detection*. Results showed a positive effect (contrary to hypothesized direction) but this effect was not significant (β = 0.047, p = 0.255). The earlier hypothesized position was that the more pressure someone was under, the less likely they were to detect a phishing threat. However, results showed different. Results implied that the more pressure a targeted individual was under, the more likely they were to detect a phishing threat. This was probably because their sense of suspicion was heightened to think that the communication was deceptive and not genuine. However, this effect was not strong enough to support the hypothesis beyond chance. Therefore, hypothesis 10a was not supported by the data.

Hypothesis 10b posited that *Pressure has a negative effect on Elaboration*. Results showed a positive effect (contrary to hypothesized direction) but this effect was not significant (β = 0.043, p = 0.41). The earlier hypothesized position was that the more pressure someone was under, the less likely they were to cognitively process a phishing message. However, results showed different. Results implied that the more pressure a targeted individual was under, the higher their elaboration. This was probably because their sense of urgency was heightened to the point of giving their attention to processing the phishing message. However, this effect was not strong enough to support the hypothesis beyond chance. Therefore, hypothesis 10b was not supported by the data.

The initial version of Hypothesis 11a posited that *Knowledge on Threat Domain has a positive effect on Threat Detection*; while hypothesis 11b posited that *Knowledge on Threat Domain has a positive effect on Elaboration*. However, the indicators for *Knowledge on Threat Domain* were combined with those for *Knowledge on Detection Cues* during Exploratory Factor Analysis to form one combined factor. This showed that the indicators used were essentially measuring the same latent construct. This combined construct was named as *Knowledge on Threat and Detection Cues (KTDC)*. Hypothesis 12 a-d were consolidated with hypothesis 11a-d by renaming the construct.

Hypothesis 11a therefore was reworded and it posited that *Knowledge on Threat and Detection Cues has a positive effect on Threat Detection*. Results on the analysis showed a significant positive effect (β = 0.459, p < 0.001). This means that the more knowledge a person has on the phishing threat and on detection cues, the more likely they are to detect a phishing threat. This effect was the strongest among all the tested antecedent relationships that lead to Threat Detection. Hypothesis 11a was supported by the data.

Hypothesis 11b was reworded and it posited that *Knowledge on Threat and Detection Cues has a positive effect on Elaboration*. Results on the analysis showed a positive effect (in line with the hypothesized direction) but this effect was marginally significant (β = 0.13, p = 0.069). This means that the more knowledge a person has on the phishing threat and on detection cues, the more likely they are to cognitively process a phishing message. However, this effect was not strong enough to support the hypothesis beyond chance. Therefore, hypothesis 11b was not supported by the data.

Hypothesis 12 a-d were dropped after consolidating them with hypothesis 11a-d after results of the Exploratory Factor Analysis showed that they were essentially measuring the same latent construct.

Hypothesis 13 was split into two sets of hypothesis. This is because the *Knowledge on Trust Determinants* construct was found to have two dimensions during Exploratory Factor Analysis. One dimension corresponded to what could be termed as low determinants of trust. Low determinants of trust are considered ineffectual when judging the legitimacy of a message because attackers are very good at falsifying them. Such low determinants of trust include things like: consistency in logo, colors, look and feel, good grammar and spelling and personalized content. The second dimension corresponded to what were considered high determinants of trust. High determinants of trust were indicators that could be effectively used to identify phishing messages. These high determinants of trust included things like: contacting someone to confirm legitimacy of communication, evaluating web addresses, links, website encryption, certificates, registration information and also using information from security tools. Therefore hypothesis 13 was used to examine the effects of low determinants of trust while a new set of hypothesis (14 a-d) were used to examine high determinants of trust.

Therefore, Hypothesis 13a was reworded to posit that *Knowledge on Low Determinants of Trust has a positive effect on Threat Detection*. Results showed a positive effect (in line with the hypothesized direction) but this effect was not significant (β = 0.005, p = 0.93). This means that the more knowledge a person has on Low Determinants of Trust, the more likely they are to detect phishing threats. However, this effect was not strong enough to support the hypothesis beyond chance. Therefore, hypothesis 13a was not supported by the data.

Hypothesis 13b was reworded to posit that *Knowledge on Low Determinants of Trust has a positive effect on Elaboration*. Results showed a positive effect (in line with the hypothesized direction) but this effect was marginally significant (β = 0.139, p = 0.064). This means that the more knowledge a person has on Low Determinants of Trust, the more likely they are to cognitively process a phishing message. However, this effect was not strong enough to support the hypothesis beyond chance. Therefore, hypothesis 13b was not supported by the data.

Hypothesis 14a was created to posit that *Knowledge on High Determinants of Trust has a positive effect on Threat Detection*. Results showed a positive effect (in line with the hypothesized direction) but this effect was not significant (β = 0.086, p = 0.139). This means that the more knowledge a person has on High Determinants of Trust, the more likely they are to detect phishing threats. However, this effect was not strong enough to support the hypothesis beyond chance. Therefore, hypothesis 14a was not supported by the data.

Hypothesis 14b was created to posit that *Knowledge on High Determinants of Trust has a positive effect on Elaboration*. Results showed a positive effect (in line with the hypothesized direction) but this effect was not significant (β = 0.002, p = 0.977). This means that the more knowledge a person has on High Determinants of Trust, the more likely they are to cognitively process a phishing message. However, this effect was not strong enough to support the hypothesis beyond chance. Therefore, hypothesis 14b was not supported by the data.

5.6. Mediation

In order to understand the model better, mediation analysis was conducted. Hypothesis 3b, 4c, 4d, 5c, 5d, 6c, 6d, 7c, 7d, 8c, 8d, 9c, 9d, 10c, 10d, 11c, 11d, 12c, 12d, 13c, 13d, 14c and 14d were used to evaluate the mediating effects of Threat Detection and Elaboration on Phishing Susceptibility for each antecedent construct. A fully connected model was used and the mediation analysis was run on IBM AMOS version 23. Results showed that mediation existed only for hypothesis 3b, 4d, 5c, 6c and 11c.

The results of the mediation analysis are summarized in Table 11.

Hypothesis 3b posited that *Threat Detection mediates the effect Elaboration has on Phishing Susceptibility*. Results showed that Threat Detection *fully mediates* the effect Elaboration has on Phishing Susceptibility. This finding shows that Threat Detection fully explains the effect Elaboration has on Phishing Susceptibility. This means that people who demonstrate high levels of Elaboration are less susceptible to phishing attacks because they are more likely to detect the threat.

Hypothesis 4d posited that *Elaboration mediates the effect Quality of Argument has on Phishing Susceptibility*. Results showed that Elaboration *indirectly mediates* the effect Quality of Argument has on Phishing Susceptibility. This is because the direct effects between Quality of Argument and Phishing Susceptibility (with and without Elaboration as a mediator) were not significant. Only the indirect effect through Elaboration as a mediator was significant. This means that Elaboration indirectly explains the effect Quality of Argument has on Phishing Susceptibility.

Hypothesis 5c posited that *Threat Detection mediates the effect Persuasive Cues have on Phishing Susceptibility*. Results showed that Threat Detection *indirectly mediates* the effect Persuasive Cues have on Phishing Susceptibility. This is because the direct effects between Persuasive Cues and Phishing Susceptibility (with and without Threat Detection as a mediator) were not significant. Only the indirect effect through Threat Detection as a mediator was significant. This means that Threat Detection indirectly explains the effect Persuasive Cues have on Phishing Susceptibility.

Hypothesis 6c posited that *Threat Detection mediates the effect Involvement has on Phishing Susceptibility*. Results showed that Threat Detection *fully mediates* the effect Involvement has on Phishing Susceptibility. This is because the direct effects between Involvement and Phishing Susceptibility became insignificant once Threat Detection was introduced as a mediator. This means that people who are highly involved in a subject matter touched on in a phishing message are less susceptible to phishing attacks because they are likely to detect the threat.

Hypothesis 11c posited that *Threat Detection mediates the effect Knowledge on Threat and Detection Cues have on Phishing Susceptibility*. Results showed that Threat Detection *indirectly mediates* the effect Knowledge on Threat and Detection Cues have on Phishing Susceptibility. This is because the direct effects between Knowledge on Threat and Detection Cues and Phishing Susceptibility (with and without Threat Detection as a mediator) were not significant. Only the indirect effect through Threat Detection as a mediator was significant. This means that Threat Detection indirectly explains the effect Knowledge on Threat and Detection Cues have on Phishing Susceptibility.

Table 11
Analysis of Mediation through Threat Detection and Elaboration.

	Path	Direct to DV without mediator	Direct to DV with mediator	Standardized Indirect Effect	Conclusion
H3b	ELAB > TD > DV	-0.168^{*}	−0.133 (ns)	-0.051^{**}	Threat Detection fully mediates the effect Elaboration has on Phishing Susceptibility.
H4c	QA > TD > DV	0.018 (ns)	0.157 (ns)	−0.027(ns)	No Mediation
H4d	QA > ELAB > DV	0.018 (ns)	0.158 (ns)	-0.118^{*}	Elaboration indirectly mediates the effect Quality of Argument has on Phishing Susceptibility.
H5c	PC > TD > DV	0.098 (ns)	0.027 (ns)	0.059^{**}	Threat Detection indirectly mediates the effect Peripheral Cues have on Phishing Susceptibility
H5d	PC > ELAB > DV	0.098 (ns)	0.027 (ns)	0.016 (ns)	No Mediation
H6c	INV > TD > DV	0.198^{*}	0.113 (ns)	0.063^{**}	Threat Detection fully mediates the effect Involvement has on Phishing Susceptibility
H6d	INV > ELAB > DV	0.198^{*}	0.114 (ns)	0.01 (ns)	No Mediation
H7c	RES > TD > DV	−0.061 (ns)	−0.102 (ns)	0.004 (ns)	No Mediation
H7d	RES > ELAB > DV	−0.061 (ns)	−0.102 (ns)	−0.015 (ns)	No Mediation
H8c	DIST > TD > DV	–	–	–	Dropped at CFA
H8d	DIST > ELAB > DV	–	–	–	Dropped at CFA
H9c	EM > TD > DV	–	–	–	Dropped at EFA
H9d	EM > ELAB > DV	–	–	–	Dropped at EFA
H10c	PRES > TD > DV	−0.09 (ns)	−0.052 (ns)	−0.014 (ns)	No Mediation
H10d	PRES > ELAB > DV	−0.09 (ns)	−0.052 (ns)	0.197 (ns)	No Mediation
H11c	KTDC > TD > DV	−0.105 (ns)	0.068 (ns)	-0.139^{**}	Threat Detection indirectly mediates the effect Knowledge on Threat Domain and Detection Cues has on Phishing Susceptibility
H11d	KTDC > ELAB > DV	−0.105 (ns)	0.068 (ns)	−0.026 (ns)	No Mediation
H12c	KDC > TD > DV	–	–	–	Merged with KTD at EFA
H12d	KDC > ELAB > DV	–	–	–	Merged with KTD at EFA
H13c	KDT_LOW > TD > DV	0.141 (ns)	0.167 (ns)	−0.002 (ns)	No Mediation
H13d	KDT_LOW > ELAB > DV	0.141 (ns)	0.168 (ns)	−0.026 (ns)	No Mediation
H14c	KDT_HIGH > TD > DV	-0.224^{*}	-0.214^{*}	−0.026 (ns)	No Mediation
H14d	KDT_HIGH > ELAB > DV	-0.224^{*}	-0.197^{*}	0 (ns)	No Mediation

Notes: *** p-value < 0.001; ** p-value < 0.010; * p-value < 0.050; ns = not significant.

5.7. Moderation

Work by Downs et al. (2007) showed that people who answered knowledge questions were significantly less likely to fall for phishing. Therefore, in addition to examining the relationships between Knowledge and Phishing Susceptibility covered by Hypothesis 11, 12, 13 and 14; this study also examined the impact that the level of knowledge had on the overall model. It was expected that some relationships in the model would differ based on a person's level of knowledge. Multi-group moderation analysis was done in an exploratory manner in order to assess whether the hypothesized effects differed based on the level of knowledge on phishing a person had.

The Knowledge Quiz section of the questionnaire was used as a grouping variable to section the dataset into two groups. The use of a knowledge quiz to assess the knowledge level of participants was borrowed and adapted from Downs et al. (2007). One point was awarded for every correct answer and therefore the maximum score for the quiz was 6. Those who scored 3 or less were grouped into a “Low Knowledge” group and those who scored 4 and above were grouped into a “High Knowledge” group. The Low Knowledge group had 125 of the 192 cases while the High Knowledge group had the remaining 67 cases.

Group differences were analyzed using Gaskin (2016) Stats Tools Package. Each group's regression weights table and the ‘Critical Ratios for Differences between Parameters’ matrices were used in the analysis. The results showed that the level of knowledge significantly moderated 5 hypothesized relationships.

The results of the moderation analysis are outlined in Table 12.

The first significant variance was on the effect of Threat Detection on Phishing Susceptibility. The effect of Threat Detection was weaker for those with high knowledge than those with low knowledge.

The second significant variance was that Elaboration had a negative effect on Phishing Susceptibility for those with low knowledge but it had a positive effect for those with high knowledge.

The third significant variance was that Quality of Argument had a positive effect on Threat Detection for those with low knowledge but it had a negative effect for those with high knowledge.

The fourth significant variance was that Peripheral Cues had a negative effect on Elaboration for those with low knowledge but had a positive effect for those with high knowledge.

The last significant difference was that Low Determinants of Trust had a positive effect on Threat Detection for those with low knowledge but had a negative effect for those with high knowledge.

6. Discussion

6.1. Discussion of findings

This study has presented a theoretical model for determining phishing susceptibility grounded on the Elaboration Likelihood Model (ELM). Few studies in social engineering research have been grounded in theory (Luo et al., 2013; Tetri & Vuorinen, 2013; Wang et al., 2012). This research aimed at addressing this gap. ELM was chosen to provide a different perspective compared to a previous study by Luo et al. (2013) that used the Heuristic-Systematic Model (HSM) and that by Wang et al. (2012) that used the Theory of Deception. More constructs from ELM were tested than in these previous studies. Furthermore, this study also examined the mediating effect that Threat Detection and Elaboration have on phishing susceptibility.

This model was empirically tested with data from 192 cases collected through direct observations of phishing susceptibility and also from self-reported questionnaires. Phishing susceptibility was determined by observing actual behaviour of a university population that was targeted with a staged phishing attack in a naturalistic field study. Other researchers (Anandpara et al., 2007; Dhamija et al., 2006) have found naturalistic field studies to provide results with higher validity compared to the use of lab studies and phishing knowledge tests.

A total of 241 users were observed to have opened the phishing email and are called the “active participants”. Although this was a 5.37% representation of the targeted population, this was comparable to a previous study by Mohebzada et al. (2012, pp. 249–254) which had a success rates of 8.74% and subsequently 2.05% in a second round. In addition, Vishwanath et al. (2011) point out that phishing exercises are often plagued by many challenges which result in low success rates and difficulties in collecting data from the targeted sample. It should also be noted that even though the success rate may be considered low, it only takes a few insiders to effectively compromise an information system.

Table 12
Moderating effect of Knowledge.

	Path	Low Knowledge		High Knowledge		z-score	Conclusion
		Estimate	P	Estimate	P		
H1	TD - > DV	−0.109	0.000	−0.200	0.000	2.18**	Effect weaker for those with High Knowledge
H2	ELAB - > DV	−0.039	0.202	0.062	0.043	−2.336**	Negative effect for those with Low Knowledge; Positive effect for those with High Knowledge
H3	ELAB - > TD	0.106	0.150	0.303	0.002	−1.588	Invariant
H4a	QA - > TD	0.194	0.023	−0.059	0.606	1.77*	Positive effect for those with Low Knowledge; Negative effect for those with High Knowledge
H4b	QA - > ELAB	0.604	0.000	0.726	0.000	−0.862	Invariant
H5a	PC - > TD	−0.268	0.000	−0.083	0.364	−1.537	Invariant
H5b	PC - > ELAB	−0.170	0.069	0.077	0.492	−1.691*	Negative effect for those with Low Knowledge; Positive effect for those with High Knowledge
H6a	INV - > TD	−0.280	0.000	−0.164	0.074	−0.997	Invariant
H6b	INV - > ELAB	−0.062	0.473	−0.050	0.658	−0.087	Invariant
H7a	RES - > TD	−0.024	0.763	−0.040	0.707	0.122	Invariant
H7b	RES - > ELAB	0.110	0.253	0.065	0.624	0.280	Invariant
H10a	PRES - > TD	0.071	0.344	0.158	0.139	−0.670	Invariant
H10b	PRES - > ELAB	0.074	0.418	0.064	0.629	0.062	Invariant
H11a	KTDC - > TD	0.484	0.000	0.630	0.000	−1.099	Invariant
H11b	KTDC - > ELAB	0.127	0.226	0.115	0.350	0.080	Invariant
H13a	KDT_LOW - > TD	0.157	0.118	−0.230	0.048	2.522**	Positive effect for those with Low Knowledge; Negative effect for those with High Knowledge
H13b	KDT_LOW - > ELAB	0.255	0.034	0.029	0.841	1.214	Invariant
H14a	KDT_HIGH - > TD	0.062	0.431	0.132	0.155	−0.579	Invariant
H14b	KDT_HIGH - > ELAB	−0.008	0.934	0.035	0.756	−0.291	Invariant

Notes: *** p-value < 0.001; ** p-value < 0.010; * p-value < 0.050.

Results showed that 31.12% of the active participants clicked the phishing link and 88% of those who clicked the link went ahead to submit their passwords on the phishing webpage.

The model of phishing susceptibility was found to have excellent Goodness-of-Fit. It was able to account for 69.5% of an individual's Threat Detection, 50.8% of Elaboration (cognitive processing) and was able to predict 28% of an individual's Phishing Susceptibility. This was better than the model by Wang et al. (2012) that explained 16.4% of the variance in the likelihood to respond to phishing and 11.9% in thought processing. This study's model was also better than that by Vishwanath et al. (2011) which accounted for 22% of the variance in cognitive processing.

The first round of Hypothesis testing examined 25 relationships based on the conceptual model. Of these, 6 were supported because their hypothesized effects were found to be significant. However, 6 hypothesis were not tested because their associated constructs (namely Distractions and Emotions) were dropped during factor analysis. It should be noted that 13 hypothesis were not supported by the data because their hypothesized effects were not found to be significant.

The second round of Hypothesis testing examined the mediating effects of Threat Detection and Elaboration on Phishing Susceptibility for each antecedent construct. Evidence of mediation was found for 5 of the 23 relationships tested. Threat Detection was found to fully mediate the effect Elaboration has on Phishing Susceptibility and also to fully mediate the effect Involvement had on Phishing Susceptibility. Threat Detection also was found to indirectly mediate the effect Peripheral Cues have on Phishing Susceptibility and also to indirectly mediate the effect Knowledge on Threat Domain and Detection Cues have on Phishing Susceptibility. Elaboration was found to indirectly mediate the effect Peripheral Cues have on Phishing Susceptibility.

The last round of analysis examined the moderating effect a person's level of knowledge had on the overall model. This involved dividing the dataset into two groups based on level of knowledge (low knowledge group and high knowledge group) and running a multi-group moderation analysis. Results showed significant differences for 5 of the hypothesized relationships depending on the level of knowledge an individual had. These relationships were between: Threat Detection and Phishing Susceptibility; Elaboration and Phishing Susceptibility, Quality of Argument and Threat Detection; Peripheral Cues and Elaboration and Knowledge on Low Determinants of Trust and Threat Detection.

The model also showed that 3 out of 12 control variables had a significant effect on the dependent variables. These 3 control variables that had a significant effect were: Role at the University, Email Load and Email Responsiveness. They were retained in the model in order to account for their effects with respect to the explicitly hypothesized relationships.

6.2. Insights on phishing susceptibility

The overall model analysis gives important insights in the understanding of what makes people susceptible to phishing attacks.

Threat detection accounts for the strongest effect with regards to reducing phishing susceptibility as compared to cognitive processing (Elaboration). The greater a person's ability to detect a phishing threat, the lower their susceptibility to phishing threats.

Cognitive processing of phishing messages was neither found to directly nor significantly affect a person's susceptibility to phishing. Mediation analysis revealed that threat detection fully mediated the effect cognitive processing had on phishing susceptibility. This means that the effect high cognitive processing has in reducing phishing susceptibility was accounted for by a person's ability to detect phishing threats. Results also showed that the more a person cognitively evaluates a phishing message, the higher their ability to detect a phishing threat. Subsequently, the less susceptible they are to phishing attacks.

The antecedent construct that had the strongest effect on cognitive processing was the quality of argument. This means that phishing messages that have very convincing arguments are the most effective in encouraging people to cognitively processing them.

The increased use of persuasive cues in phishing messages was found to reduce threat detection. In addition, the more involved a person was in the subject matter communicated in a phishing attack, the less likely they are to detect the phishing attack. Mediation analysis showed that both the increased use of persuasive cues and increased involvement led to higher phishing susceptibility due to a decrease in a person's ability to detect the threat.

When considering these insights regarding the use of quality of argument and peripheral cues; attackers could devise effective strategies that deliver phishing attacks with low argument quality but high persuasive peripheral cues. Such attack strategies could be used to discourage cognitive processing and at the same time evade threat detection.

The antecedent construct that had the highest effect on a person's ability to detect threats was their knowledge on phishing threats and on phishing detection cues. The more knowledge a person had, the more likely they would be to detect a phishing threat. Mediation analysis showed that threat detection indirectly accounted for the reason why people with higher levels of knowledge are less susceptible to phishing attacks.

Moderation analysis showed that the model worked differently depending on an individual's level of knowledge. These differences were with regards to the effects of: threat detection on phishing susceptibility; cognitive processing on phishing susceptibility; quality of argument on threat detection; persuasive cues on cognitive processing and the effect the knowledge on low determinants of trust has on threat detection.

The effect that increased threat detection had on reducing phishing susceptibility was weaker for those with high knowledge. In addition, it was found that when people with low levels of knowledge engaged in increased cognitive processing, they experienced reduced phishing susceptibility.

In addition, it was also found that when people with low knowledge encountered a phishing message with high quality of argument, they were more likely to detect the threat. However, when people with high levels of knowledge encountered the phishing message with high quality of argument, they were less likely to detect the threat.

With regards to persuasive cues, it was found that when people with low levels of knowledge encountered phishing messages that had increased use of peripheral cues, they were less likely to expend effort to cognitively process the phishing message. However when people with high levels of knowledge encountered phishing messages that had increased use of peripheral cues, they were more likely to expend cognitive effort to process the message.

With regards to knowledge on low determinants of trust (such as good look and feel, no grammatical errors), it was found that people with low levels of knowledge relied on them more for threat detection than people of high levels of knowledge did.

6.3. Implications on practice

These results show that the best way to stop users from falling for phishing is by equipping them with the knowledge and skills to effectively detect phishing attacks. Effective detection of phishing attacks could even preclude the active interaction with phishing messages. Users who do not interact with phishing artifacts in the first place are better protected than those who actively interact with them and thereafter need additional rationale to evade the attack.

Users need to be specifically trained on “high determinants of trust” and not just ‘general’ phishing knowledge. These high determinants of trust include: taking time to contact a help desk to confirm legitimacy of communication, using specific criteria to evaluate email headers, web addresses, hyperlinks, encryption, certificates, address registration information, among others. They could also be equipped with phishing detection tools that implement these trust criteria and support them in phishing detection. However, it is important that the users are also trained on how to use the information provided by the tools and most importantly that they do not ignore warnings.

The implication of this finding is that users have to invest time and effort to think. Without cognitive processing the users are still susceptible to phishing because it is the primary element that leads to phishing detection. Therefore, users need to be warned against responding to messages without much thought. They should not respond to messages due to emotional impulses or pressure. Mechanisms should be provided to allow them to 'fact check' messages meant to elicit such responses from them.

6.4. Limitations of the study

A key limitation of this study is that it was only able to study 5.37% of the sample and therefore was unable to capture the majority of the targeted participants. Vishwanath et al. (2011) point out that phishing studies are often plagued by many challenges that result in such low response rates.

Future studies should consider ways to increase interaction with phishing stimuli and participation from targeted users. The model analysis and results could be markedly different with data from a larger representation of targeted participants. It could be that those who did not interact with the phishing email had reasons to suspect that it was a phishing attack. The study could benefit from analyzing such data. However, care should still be taken not to contaminate the analysis with data from people who did not suspect phishing, for example, users who did not even access their emails in the first place.

7. Conclusions

The model presented in this study is robust and has brought forth key insights in explaining the effects that cognitive processing and threat detection have on phishing susceptibility. The effects that various antecedent factors to these constructs have has been analyzed using Hypothesis testing, mediation analysis and multi-group moderation. The model was empirically tested using data from 192 cases collected through direct observations of phishing susceptibility and also from self-reported questionnaires.

This study has outlined a naturalistic field study research methodology that is useful when guiding future work in order to deliver research with high ecological validity. The study has also show cased various tools, variables and measurement items that can guide future empirical studies in this area.

Further work can be done to extend the model in order to better account for the variance in phishing susceptibility, threat detection and cognitive process. The model should also be empirically tested with a larger dataset and also with data from populations in more diverse contexts to allow for better generalizability of findings.

Threat detection was found to have the strongest effect in reducing phishing susceptibility. Organizations should invest in mitigation measures that support users in detecting phishing threats. It is expected that phishing attack strategies could aim at discouraging cognitive processing in order to evade threat detection. Organizations should be aware of such strategies and intentionally implement measures that encourage their users to carefully process messages before responding to them. In addition, organizations should sensitize their users on phishing threats and also equip them with skills which enable them to effectively use detection cues to identify phishing attacks.

Funding

This work was supported by the German Academic Exchange Service or DAAD (In-Country/In-Region Scholarships for Postgraduates).

References

Anandpara, V., Dingman, A., Jakobsson, M., Liu, D., & Roinestad, H. (2007). Phishing IQ tests measure fear, not ability. *Presented at the international conference on financial cryptography and data security* (pp. 362–366). Trinidad and Tobago: Springer Berlin Heidelberg.

Angst, C. M., & Agarwal, R. (2009). Adoption of electronic Health records in the presence of privacy concerns: The elaboration likelihood model and individual persuasion. *MIS Quarterly, 33*(2), 339–370.

APWG, A.-P. W. G. (2016). *Phishing activity trends report: 1st quarter 2016*Anti-Phishing Working Group. Retrieved from https://docs.apwg.org/reports/apwg_trends_report_q1_2016.pdf.

APWG, A.-P. W. G. (2017). *Phishing activity trends report: 4th quarter 2016*Anti-Phishing Working Group. Retrieved from https://docs.apwg.org/reports/apwg_trends_report_q4_2016.pdf.

APWG, A.-P. W. G. (2018). *Phishing activity trends report: 3rd quarter 2017*Anti-Phishing Working Group. Retrieved from http://docs.apwg.org/reports/apwg_trends_report_q3_2017.pdf.

Arachchilage, N. A. G., & Love, S. (2013). A game design framework for avoiding phishing attacks. *Computers in Human Behavior, 29*, 706–714.

Bakhshi, T., Papadaki, M., & Furnell, S. (2009). Social engineering: Assessing vulnerabilities in practice. *Information Management & Computer Security, 17*(1), 53–63. https://doi.org/10.1108/09685220910944768.

Barth, B. (2016, August 1). *Don't be like "Mike": Authorities arrest mastermind of $60M online scam operation.* SC Magazine.

BBC News. (2016, August 1). *Online fraud: Top Nigerian scammer arrested.* BBC News. Retrieved from http://www.bbc.com/news/world-africa-36939751.

Bhattacherjee, A., & Sanford, C. (2006). Influence processes for information technology acceptance: An elaboration likelihood model. *MIS Quarterly, 30*(4), 805–825.

Brehm, J. W. (1966). *A theory of psychological reactance.* New York: Academic Press.

Bulgurcu, B., Cavusoglu, H., & Benbasat, I. (2010). Information security policy compliance: An empirical study of rationality-based beliefs and information security awareness. *MIS Quarterly, 34*(3), 523–548.

Buller, D. B., & Burgoon, J. K. (1996). Interpersonal deception theory. *Communication Theory, 6*(3), 203–242.

CERT, I. T. T. (2013). *Unintentional insider threats: A foundational study.* Software Engineering Institute, Carnegie Mellon University. Retrieved from http://resources.sei.cmu.edu/asset_files/TechnicalNote/2013_004_001_58748.pdf.

Chaiken, S. (1980). Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of Personality and Social Psychology, 39*(5), 752–766.

Cialdini, R. B. (2001). *Influence: Science and practice* (4th ed.). Boston MA: Allyn & Bacon.

Cimpanu, C. (2016, April 28). *Anonymous hackers leak 1 TB of documents from Kenya's ministry of foreign affairs.* Retrieved http://news.softpedia.com/news/anonymous-hackers-leak-1tb-of-documents-from-kenya-s-ministry-of-foreign-affairs-503518.shtml, Accessed date: 12 September 2016.

Curran, P. J., West, S. G., & Finch, J. F. (1996). The robustness of test Statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods, 1*(1), 16–29.

Cyveillance (2015). *The cost of phishing: Understanding the true cost dynamics behind phishing attacks.* Cyveillance.

Dhamija, R., Tygar, J. D., & Hearst, M. (2006). Why phishing works. *Proceedings of the SIGCHI conference on Human Factors in computing systems* (pp. 581–590). ACM.

Dodge, R. C., Carver, C., & Ferguson, A. J. (2007). Phishing for user security awareness. *Computers & Security, 26*, 73–80.

Downs, J. S., Holbrook, M., & Cranor, L. F. (2006). Decision strategies and susceptibility to phishing. *Proceedings of the second symposium on Usable privacy and security* (pp. 79–90). ACM.

Downs, J. S., Holbrook, M., & Cranor, L. F. (2007). Behavioral response to phishing risk. *Proceedings of the anti-phishing working groups 2nd annual eCrime researchers summit* (pp. 37–44). ACM.

Eagly, A. H., & Chaiken, S. (1993). *The psychology of attitudes.* Harcourt Brace Jovanovich College Publishers.

Festinger, L. (1957). Cognitive dissonance theory. *Primary prevention of HIV/AIDS: Psychological approaches*. Newbury Park, California: SAGE Publications.

Finn, P., & Jakobsson, M. (2007). Designing and conducting phishing experiments. *IEEE Technology and Society Magazine, Special Issue on Usability and Security, 26*(1), 46–58.

Fire Eye (2017, January). *APT28: At the center of the storm.* Retrieved from https://www.fireeye.com/current-threats/threat-intelligence-reports.html.

Fogg, B. J., Marshall, J., Laraki, O., Osipovich, A., Varma, C., Fang, N., et al. (2001). What makes web sites credible? A report on a large quantitative study. *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 61–68). ACM.

Fornell, C., & Larcker, D. F. (1981). Structural equation models with unobservable variables and measurement error: Algebra and statistics. *Journal of Marketing Research,* 382–388.

Friedman, B., Hurley, D., Howe, D. C., Felten, E., & Nissenbaum, H. (2002). Users' conceptions of web security: A comparative study. *Extended abstracts on human factors in computing systems* (pp. 746–747). ACM.

Garera, S., Provos, N., Chew, M., & Rubin, A. D. (2007). A framework for detection and measurement of phishing attacks. *Proceedings of the 2007 ACM workshop on Recurring malcode*. ACM.

Gaskin, J. (2016). *Stats tools package.* Retrieved from http://statwiki.kolobkreations.com.

Grazioli, S. (2004). Where did they go wrong? An analysis of the failure of knowledgeable internet consumers to detect deception over the internet. *Group Decision and Negotiation, 13*(2), 149–172.

Grazioli, S., & Jarvenpaa, S. (2001). Tactics used against consumers as victims of internet deception. *AMCIS 2001 proceedings* (pp. 810–815). Boston.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2009). *Multivariate data analysis* (7th ed.). .

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1–55.

Huber, M., Kowalski, S., Nohlberg, M., & Tjoa, S. (2009). Towards automating social engineering using social networking sites. *Presented at the computational science and engineering: Vol. 3,* (pp. 117–124). IEEE.

Jagatic, T., Johnson, N., Jakobsson, M., & Menczer, F. (2007). Social phishing. *Communications of the ACM, 50*(10), 94–100.

Jakobsson, M. (2005). Modeling and preventing phishing attacks. *Financial Cryptography, 5*.

Jakobsson, M., & Ratkiewicz, J. (2006). Designing ethical phishing experiments: A study of (ROT13) rOnl query features. *Proceedings of the 15th international conference on world wide web*. ACM.

Jakobsson, M., Tsow, A., Shah, A., Blevis, E., & Lim, Y.-K. (2007). What instills trust? A qualitative study of phishing. *Financial cryptography and data security* (pp. 356–361). Springer Berlin Heidelberg.

James, L. (2005). *Phishing exposed.* Syngress.

Johnson, P. E., Grazioli, S., Jamal, K., & Berryman, R. G. (2001). Detecting deception: Adversarial problem solving in a low base-rate world. *Cognitive Science, 25*(3), 355–392.

Johnson, P. E., Grazioli, S., Jamal, K., & Zualkernan, I. A. (1992). Success and failure in expert reasoning. *Organizational Behavior and Human Decision Processes, 53*(2), 173–203.

Karakasiliotis, Furnell, S. M., & Papadaki, M. (2006). Assessing end-user awareness of social engineering and phishing. *Proceedings of 7th Australian information warfare and security conference*. Perth Western Australia: Edith Cowan University.

Kline, R. B. (2013). Exploratory and confirmatory factor analysis. In Y. Petscher, & C. Schatsschneider (Eds.). *Applied quantitative analysis in the social sciences* (pp. 171–207). New York: Routledge.

Kumaraguru, P., Cranshaw, J., Acquisti, A., Cranor, L., Hong, J., Blair, M. A., et al. (2009). School of phish: A real-word evaluation of anti-phishing training. *Proceedings of the 5th symposium on usable privacy and security (SOUPS)*. Mountain View, CA, USA: ACM.

Kumaraguru, P., Rhee, Y., Acquisti, A., Cranor, L. F., Hong, J., & Nunge, E. (2007a). Protecting people from phishing: The design and evaluation of an embedded training email system. *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 905–914). ACM.

Kumaraguru, P., Rhee, Y., Sheng, S., Hasan, S., Acquisti, A., Cranor, L. F., et al. (2007b). Getting users to pay attention to anti-phishing education: Evaluation of retention and transfer. *Proceedings of the anti-phishing working groups 2nd annual eCrime researchers summit* (pp. 70–81). ACM.

Kumaraguru, P., Sheng, S., Acquisti, A., Cranor, L. F., & Hong, J. (2007c). Teaching Johnny not to fall for phish. *CM Transactions on Internet Technology (TOIT), 10*(2), 7.

Kumaraguru, P., Sheng, S., Acquisti, A., Cranor, L., & Hong, J. (2008). Lessons from a real world evaluation of anti-phishing training. *Presented at the eCrime researcher's summit*. Anti-Phishing Working Group (APWG).

LaRose, R., Rifon, N. J., & Enbody, R. (2008). Promoting personal responsibility for internet safety. *Communications of the ACM, 51*(3), 71–76.

Liang, H., & Xue, Y. (2009). Avoidance of information technology threats: A theoretical perspective. *MIS Quarterly, 33*(1), 71–90.

Liang, H., & Xue, Y. (2010). Understanding security behaviors in personal computer usage: A threat avoidance perspective. *Journal of the Association for Information Systems (JAIS), 11*(7), 394–413.

Luo, X. (Robert), Zhang, W., Burd, S., & Seazzu, A. (2013). Investigating phishing victimization with the heuristic-systematic model: A theoretical framework and an exploration. *Computers & Security, 38*, 28–38.

Mitnick, K. D., & Simon, W. L. (2002). *The art of deception: Controlling the human element of security.* John Wiley & Sons.

Mohebzada, J. G., El Zarka, A., BHojani, A. H., & Darwish, A. (2012). Phishing in a university community: Two large scale phishing experiments. *Presented at the international conference on innovations in information technology (IIT)* (pp. 249–254). IEEE.

Nunnally, J., & Bernstein, I. (1994). *Psychometric theory* (2nd ed.). New York: McGraw Hill.

Obulutsa, G. (2016, April 28). *Hackers leak stolen Kenyan foreign ministry documents.* Retrieved http://www.reuters.com/article/us-cyber-kenya-idUSKCN0XP2K5, Accessed date: 12 September 2016.

Parsons, M. H. (1974). What happened at Hawthorne? *Science, 183*(4128), 922–932.

Petty, R. E. (1994). Two routes to persuasion: State of the art. In G. d'Ydewalle, P. Eelen, & P. Berteleson (Vol. Eds.), *International perspectives on psychological science: Vol. 2*, (pp. 229–247). Hillsdale, NJ: Erlbaum.

Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. *Advances in Experimental Social Psychology, 19*.

Petty, R. E., & Wegener, D. T. (1998). Attitude change: Multiple roles for persuasion variables. *Handbook of social psychology* (pp. 323–390). (4th ed.). New York: McGraw Hill.

Petty, R. E., & Wegener, D. T. (1999). The elaboration likelihood model: Current status and controversies. In S. Chaiken, & Y. Trope (Eds.). *Dual-process theories in social psychology* (pp. 37–72). New York, NY, US: Guilford Press.

PhishTank (2016). *PhishTank Stats.* Retrieved https://www.phishtank.com/stats.php, Accessed date: 14 October 2016.

Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: a critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*(5), 879–903.

Sheng, S., Holbrook, M., Kumaraguru, P., Cranor, L. F., & Downs, J. S. (2010). Who falls for phish? A demographic analysis of phishing susceptibility and effectiveness of interventions. *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 373–382). ACM.

Sheng, S., Magnien, B., Kumaraguru, P., Acquisti, A., Cranor, L. F., Hong, J., et al. (2007). Anti-phishing phil: The design and evaluation of a game that teaches people not to fall for phish. *Proceedings of the 3rd symposium on Usable privacy and security* (pp. 88–99). ACM.

Tetri, P., & Vuorinen, J. (2013). Dissecting social engineering. *Behaviour & Information Technology, 32*(10), 1014–1023. https://doi.org/10.1080/0144929X.2013.763860.

Tsow, A., & Jakobsson, M. (2007). *Deceit and deception: A large user study of phishing.* Indiana University.

Vishwanath, A., Herath, T., Chen, R., Wang, J., & Rao, R. H. (2011). Why do people get phished? Testing individual differences in phishing vulnerability within an integrated, information processing model. *Decision Support Systems, 51*(3), 576–586.

Wang, J., Herath, T., Chen, R., Vishwanath, A., & Rao, R. H. (2012). Phishing susceptibility: An investigation into the processing of a targeted spear phishing email. *IEEE Transactions on Professional Communication, 55*(4), 345–362.

Waqas (2016, April 28). *Anonymous leaks 1TB of data from Kenya's ministry of foreign affairs.* Retrieved https://www.hackread.com/anonymous-hacks-kenya-ministry-foreign-affairs/, Accessed date: 12 September 2016.

Workman, M. (2007). Gaining access with social engineering: An empirical study of the threat. *Information Systems Security, 16*(6), 315–331.

Workman, M. (2008a). A test of interventions for security threats from social engineering. *Information Management & Computer Security, 16*(5), 463–483.

Workman, M. (2008b). Wisecrackers: A theory-grounded investigation of phishing and pretext social engineering threats to information security. *Journal of the American Society for Information Science and Technology, 59*(4), 662–674.

The African Journal of Information Systems

A Naturalistic Methodology for Assessing Susceptibility to Social Engineering Through Phishing

Research Paper

Volume 11, Issue 3, July 2019, ISSN 1936-0282

Paula M. W. Musuva
United States International University - Africa
pmusuva@gmail.com

Christopher K. Chepken, PhD
University of Nairobi, Kenya
chepken@uonbi.ac.ke

Katherine W. Getao, PhD, EBS
Ministry of ICT, Kenya
kate.getao@gmail.com

(*Received May 2018, Accepted December 2018*)

ABSTRACT

Phishing continues to be a prevalent social engineering attack. Attacks are relatively easy to setup and can target many people at low cost. This study presents a naturalistic field experiment that can be staged by organisations to determine their exposure. This exercise provides results with high ecological validity and can give organisations the information they need to craft countermeasures to social engineering risks. The study was conducted at a university campus in Kenya where 241 valid system users, also known as "insiders," are targeted in a staged phishing experiment. The results show that 31.12% of the insiders are susceptible to phishing and 88% of them disclose passwords that grant access to attackers. This study outlines various ethical considerations that ensure such exercises do not present any actual harm. The design of data collection instruments is discussed in depth to allow organisations the opportunity to develop similar tools for routine threat assessment.

Keywords

Social Engineering, Phishing, Unintentional Insider Threat, Threat Assessment, Naturalistic Methodology, Information Security.

INTRODUCTION

Social engineering is the use of manipulation by malicious outsiders to get unsuspecting insiders to compromise an organization's information security by providing access to confidential information or protected information systems (Luo, Brody, Seazzu, & Burd, 2011). One prevalent type of social engineering is phishing. Social engineering through phishing is a type of unintentional insider threat.

The term insider is used to refer to authorized users of information systems who operate within an organization's trust boundaries. These insiders often pose as information security threats when they accidentally expose their systems to attack. This is referred to as the unintentional insider threat (CERT, 2013).

Phishing is described by the Anti-Phishing Working Group (APWG, 2018) as a criminal attack that uses deception over a technical medium in order to get users to give out their personal data, login credentials and other confidential information. The deception aims at getting the user to think that the communication is a legitimate request for their confidential data or system access. Another way to describe phishing is simply 'fishing' for data (James, 2005). This is the use of social deception (the fishing bait) with the aid of communication technologies such as apps, email or websites (the fishing rod) to compromise the security of an information system (the catch).

Background

The most common technique for delivering phishing attacks is email because it provides a way to reach large numbers of people with little effort and low cost (APWG, 2018; James, 2005; Kumaraguru, Rhee, Acquisti, et al., 2007). In addition, once an email is delivered to an insider's inbox, it is considered to have crossed the external perimeter defenses and is now inside an organization's network. This makes it a very effective way of compromising information systems from within the organization. Phishing emails are also commonly used to deliver malware onto a user's system which then harvests confidential information and automates the attack process from within the network.

Research by Verizon (2015, 2016, 2017), Fire Eye (2015, 2017) and Mandiant (2004, 2010), on recent cases of the Advanced Persistent Threat (APT) involving crimeware and cyber-espionage, show that a common technique of compromising organizations is by delivering phishing emails to targeted individuals. This phishing technique of crafting attacks to fit targeted individuals is called spear phishing. The spear phishing email is often crafted to be relevant to the recipient and also appears to come from a legitimate sender, such as a colleague or company executive, often through the use of forged e-mail addresses.

Cases of phishing attacks are still on the rise despite a long history of phishing campaigns dating back to 1995 (James, 2005). The Anti-Phishing Working Group report (APWG, 2017) reported in the fourth quarter of 2016 an increase of 65% in the number of phishing attacks compared to those reported in 2015. In addition, a trend analysis of phishing attacks since 2004 show a 5,753% increase over a 12-year period. The previous report for the first quarter of 2016 (APWG, 2016) showed a 250% increase in the number of unique phishing websites since the last quarter of 2015. PhishTank, another organization that monitors cases of phishing, reported 4.5 million phishing sites in October 2016, 42,788 of which were confirmed to be active phishing sites (PhishTank, 2016).

Research by Cyveillance (2015) on the cost of phishing shows that phishing attacks are estimated to result in losses of 5.9 billion US dollars annually. News in August 2016 (Barth, 2016; BBC News, 2016) highlighted a criminal network led by a 40 year old Nigerian man called "Mike" that had scammed individuals and companies off 60 million US dollars through email scams and phishing malware. Previous research done by Hernandez, Regalado, & Villeneuve (2015) on Nigerian scammers show consistent use of email-based social engineering to defraud businesses of millions of dollars.

Investigative reports on allegations of Russia's involvement in the 2016 elections in the United States of America also show compromise through spear-phishing emails targeted at key staff in the Democratic Party (Fire Eye, 2017). In addition, there was the 2016 attack by the group Anonymous against Kenya's

Ministry of Foreign affairs. In April 2016 Anonymous posted 1 Tera Byte (TB) of sensitive data from the ministry on the dark web. After the disclosure of the breach, Kenya's ICT Cabinet Secretary explained that the hackers succeeded in gaining access to the ministry's data through phishing. An email circulated by the head of IT dated 4th August 2015 (months before the attack) tried to alert staff on the phishing attempts being sent by people impersonating the ICT administrator (Cimpanu, 2016; Obulutsa, 2016; Waqas, 2016).

Research Problem

These recent cases of phishing attacks demonstrate that it is still an active threat to users and a growing concern for organizations today. Many organizations have focused on the use of technology without giving much attention to addressing the human factor (Luo, Brody, Seazzu, & Burd, 2011). Kevin Mitnick, one of the most renowned hackers of our time, has confessed that hackers use social engineering to exploit people since they are the weakest link even in the most secure systems (Mitnick & Simon, 2002). Mitnick points out that organizations spend a lot of money developing and implementing the best security without addressing the weak human factor in the security chain.

Organizations need a credible methodology to regularly assess their susceptibility to phishing (Dodge, Carver, & Ferguson, 2007; Ferguson, 2005; Jackson, Ferguson, & Cobb, 2005; Kumaraguru, Sheng, Acquisti, Cranor, & Hong, 2007). Results of such assessments can guide the selection and implementation of appropriate countermeasures.

This study is a response to this gap. It seeks to identify a credible methodology that information security researchers can use to assess organizational exposure to phishing threats targeted at insiders. The assessment can be done regularly and a security baseline metric can be established to routinely compare with. Assessment results can be tracked over a period of time and the effectiveness of implemented countermeasures examined to see their effectiveness in reducing insider susceptibility to social engineering attacks.

Research Question

The question that this study seeks to answer is: "How can information security researchers credibly assess the vulnerability of insiders to phishing threats?"

This study aims to present a study methodology credible in the assessment of vulnerability to phishing threats. This methodology is then employed at a university in Kenya to study its vulnerability to phishing threats. The instruments used to carry out the assessment are outlined in detail and the lessons learnt in this process are then synthesized and presented in a way that can guide practice.

LITERATURE REVIEW

A review of literature reveals three key techniques that have been used to assess vulnerability to phishing threats. One of the techniques used has been the administration of phishing knowledge tests (also known as phishing IQ tests) in questionnaires or survey instruments. Wang et al. (2012) and Vishwanath et al. (2011) used questionnaires containing images of a phishing attack that had previously been launched against a university population. They asked the respondents to indicate their likelihood to respond to the phishing email that was presented. They were not able to examine actual user responses to the phishing attack because they did not stage the attack themselves. They relied on participant self-evaluation responses to gauge phishing susceptibility. Similarly, Downs et al. (2007) presented an online questionnaire survey to 232 respondents showing images of five emails and asking them to indicate how they would respond to each email in order to gauge their phishing susceptibility. They also administered a knowledge test to gauge the participants' understanding of padlock icons and selected terminology as relates to phishing. In another related study, Tsow & Jakobsson (2007) administered an online questionnaire survey to 435 participants displaying six emails and six webpage screenshots. They asked the participants to score each screenshot, on a scale of 1-5, on how much they believed the messages to be a phishing ploy or to be genuine communication. They then used the data to evaluate both trust and deceptive tactics used in phishing scams.

A second technique commonly used in the assessment of phishing susceptibility is the conducting of lab experiments. Sheng et al. (2010), Kumaraguru, Rhee, Sheng, et al. (2007), Kumaraguru, Rhee, Acquisti, et al. (2007), Kumaraguru, Sheng, et al. (2007), Jakobsson, (2007), Jakobsson et al. (2007) and Downs et al. (2006) recruited non-expert volunteers to take part in lab experiments. Their studies involved role-play exercises staged in a lab environment where participants were asked to process a set of emails with hyperlinks. The participants were required to speak out their thought process so that the researchers could listen and identify the criteria used to distinguish phishing emails from legitimate ones. The researchers analyzed the participant feedback using what they termed as the 'think-aloud' protocol. They were then able to outline the various techniques and criteria participants used to identify phishing emails. These studies also incorporated anti-phishing training in different variations to assess the efficacy of different training approaches as treatments to address phishing susceptibility. The use of embedded training was singled out as the most effective method of delivering anti-phishing training. Another related study by Egelman et al. (2008) used a lab experiment to present 60 participants with phishing messages and to observe their interaction with browser-based warnings. The participants were divided into four groups whereby three of the groups received browser warnings when they interacted with phishing links and the control group did not. They found that 97% of the participants were susceptible to at least one phishing attack.

A third technique used in determining the susceptibility to phishing threats is the staging of real-world phishing attacks. Luo et al. (2013) asked graduate students in an information assurance class to conduct a phishing attack targeting 105 staff and faculty in the School of Management at a southwest US university. The phishing attack was designed to imitate urgent school email communications. The pretext scenario used a survey regarding possible budget cuts affecting the targeted academic and administrative staff. A total of 38 users (36%) clicked the link and 16 (15%) disclosed their usernames and password credentials on the phishing forms. Similarly, Bakhshi et al. (2009) staged a phishing attack targeting a single department in an organization with over 2,000 users. A phishing email was sent to 152 staff requesting them to install an important software update by clicking a hyperlink to an external website. A total of 35 people (23%) clicked the hyperlink and also clicked a button marked 'Proceed' to install the software update. In other related studies, Kumaraguru et al. (2009) and Kumaraguru et al. (2008) staged phishing attacks targeting a university population and a large corporation respectively. They purchased domains, set up real websites and delivered phishing emails to targeted participants. They found that 90% of participants who are vulnerable to phishing will click links within 8 hours of a phishing attack being delivered to them. They also designed their experiments to assess the effectiveness of various training options in reducing susceptibility to phishing attacks. In another study, Jagatic et al. (2007) crawled a popular social media site to extract profile information that was later used to customize phishing attacks. Results revealed a 72% success rate when the information was used to customize phishing attacks. Other studies by Dodge et al. (2007) and Jackson, Ferguson, & Cobb (2005) involved staging phishing attacks targeted at students studying at a Military Academy in West Point, Untied States. Four phishing scenarios were designed into the attack namely: clicking of hyperlinks, opening of attachments, submitting sensitive information to online forms and the installation of downloaded applications. Results showed that 29% of those targeted clicked phishing links, 47% opened phishing attachments and 45% submitted sensitive information on a staged phishing website. The researchers were unable to stage the download and installation of a questionable application due to security and privacy concerns.

Each of these techniques has its own set of strengths and limitations as discussed hereafter. The use of phishing knowledge tests has a number of strengths. The technique is easily incorporated into questionnaires and survey instruments that can be distributed to many participants to collect data using uniform measures and verified scales. This makes it a very cost effective method of assessing phishing susceptibility and does not require the setup of technical infrastructure for phishing. The questionnaires can also be used to measure non-observable constructs associated with phishing susceptibility such as; intentions, attitudes and perceptions. The use of knowledge tests is associated with notable limitations making them unsuitable. Anandpara, Dingman, Jakobsson, Liu, & Roinestad (2007) demonstrated that these tests do not measure capabilities and skills in detecting phishing attacks. In fact, scoring highly in the tests may give participants a false sense of confidence that they are not susceptible to the threat. Additionally, knowledge tests require participants to self-report and may have elements requiring participants to recall their actions or thought processes from events that took place in the past. This can introduce measurement bias because people are known to assess themselves more favorably than they would act in practice. In addition, people tend to forget and may make up facts to fill in gaps in their recollection of past events. Participants may also respond in ways that are considered ‘acceptable’ to the researchers because they know they are under study. The Hawthorne effect (Parsons, 1974) explains that study participants are known to alter their behaviour due to the awareness of being studied. This leads to contamination of results. Additionally, these knowledge tests often use static content (such as screenshots) to illicit participant evaluations. Such static content is devoid of many interactive security indicators that would be available to users in real-life settings to identify phishing attacks.

The use of lab experiments in phishing assessments has its strengths. Researchers have been known to use ‘think-aloud’ protocols and observation techniques that provide rich insights in the participants’ thought and decision-making processes. The data collected through such protocols allows for quantitative and qualitative analyses.

Conversely, lab experiments require considerable technical expertise to simulate real-world phishing attacks in a lab setting. In addition, resource constraints of a lab setup can make it difficult to engage many study participants at the same time. Past studies by Kumaraguru, Rhee, Sheng, et al. (2007), Kumaraguru, Rhee, Acquisti, et al. (2007), Kumaraguru, Sheng, et al. (2007) engaged totals of 49, 30 and 28 participants respectively, with each participant being interactively engaged during data collection. Richer engagement in study protocols could also mean more time and effort during data collection. It can also be argued that simulated environments are not comparable to real attacks. They may create a false sense of security in participants because they are not exposed the real consequences of a phishing attack. Consequently, participants may be more willing to take actions that they would otherwise not take when under a real attack as was observed by Downs et al. (2007). Furthermore, participants know they are being studied and in many cases they are primed to look out for the threat (Dhamija, Tygar, & Hearst, 2006). This heightens their awareness and alters their behaviour contrary to what would have been the case in their normal day-to-day activities. This behaviour modification contaminates the results of the study and compromises the validity and reliability of results. In addition, the selection of participants for lab studies may also introduce bias. Such recruitment often requires participants to volunteer to take part in the study and may also use convenience samples. Consequently, there could be unique characteristics about the type of participants who take part - meaning they are not a good representation of the general population as noted by Kumaraguru, Sheng, et al. (2007) and Downs et al. (2007). This threatens the ecological validity of the study and makes it harder to generalize the findings to real-life settings and to more diverse populations.

The use of a naturalistic studies incorporating staged attacks that mimic real-world threats is arguably the recommended method of assessing susceptibility to phishing threats. Finn & Jakobsson (2007) argue that they are more effective than lab studies or knowledge tests. This is because naturalistic studies seek to observe actual behaviour in its normal context. The insiders are not made aware of the ongoing study and are expected to operate as they normally would in the absence of the study. This protects against the Hawthorne effect. In addition, Huber et al. (2009), Kumaraguru et al. (2009) and Workman (2007, 2008a) point out that such naturalistic studies have high ecological validity. Brewer & Crano (2014) explain that ecological validity is associated with studies whose settings approximate the real-world scenarios and what is everyday life for the wider population. High ecological validity enables results to be generalized to wider populations with similar real-world settings. In addition, the infrastructure required to stage phishing attacks is now readily accessible and fairly easy to setup as demonstrated by graduate students in the study by Luo et al. (2013). Researchers can purchase domains, setup web servers and carry out mass mailing to target large populations in a straightforward manner. The phishing instruments can also include active scripts and backend tools to collect a diverse collection of data about targeted users' online behaviours, even without alerting them. This avails rich data to researchers that allows them to build holistic user profiles. In addition, this data can be collected from many participants simultaneously.

Despite these advantages of using naturalistic field studies, Huber et al. (2009) and Kumaraguru et al. (2009) acknowledge that they are more difficult to conduct. It is difficult to get organizations willing to cooperate with the researcher to stage attacks that are as realistic, convincing and deceptive as would real attacks. In addition, such studies require approvals from research and ethical review boards which may be hard to get due to associated research risks. Many ethical review boards may be concerned by the use of deception and waiver of informed consent by participants (Finn & Jakobsson, 2007). Therefore, key to the success of such research is to identify an organization that is willing to have a naturalistic study conducted. Such an organization would give a site approval for the research on behalf of its population, with adequate oversight to ensure that there is no actual harm. Another challenge in delivering naturalistic studies is the technical expertise needed to deliver very realistic phishing attacks. The process often involves registration of domains, setting up of webservers, backend databases and phishing accounts.

Table 1 outlines a summary of the different techniques used to assess susceptibility to phishing.

Table 1: Critique of assessment techniques

Technique	Previous Studies Employing Technique	Strengths of Technique	Limitations of Technique
Knowledge / Phishing IQ Tests	• Wang et al. (2012) • Vishwanath et al. (2011) • Downs et al. (2007) • Tsow & Jakobsson (2007)	• Easy to administer in form of questionnaires or surveys. • Data can be collected in a uniform way using well-defined measures. • Cost effective method of collecting data from many participants. • Useful in measuring non-observable constructs, such as, intentions, attitudes and perceptions.	• Requires respondent to remember their past behaviour. Their memory may fail them. • Data collected subjected to self-reporting bias • People may assess themselves more (or even less) favorably than would actually act. • Results may be contaminated by Hawthorne-effects. • Not a true reflection of real-world attacks. • Many interactive features/tools not available to users for use in identifying phishing attacks.
Lab	• Sheng et al. (2010)	• Researchers can engage	• Researchers may find it hard to engage many

Technique	Previous Studies Employing Technique	Strengths of Technique	Limitations of Technique
Experiments	• Kumaraguru, Rhee, Sheng, et al. (2007) • Kumaraguru, Rhee, Acquisti, et al. (2007) • Kumaraguru, Sheng, et al. (2007) • Jakobsson, (2007) • Jakobsson et al. (2007) • Downs et al. (2006) • Egelman et al. (2008)	participants in "think-aloud" protocol technique to better understand their thought processes and behaviour. • Can allow richer data set involving qualitative and quantitative elements during study	participants at the same time. • Involves more time and effort in data collection. • Technical expertise is need to simulate a research environment that provides a set of features/tools to match real-world settings. • Not a true reflection of real-world attacks. • Participants are shielded from 'real' consequences. • Susceptible to Hawthorne-effects. • Results may not be generalizable.
Naturalistic Experiments	• Luo et al. (2013) • Bakhshi et al. (2009) • Kumaraguru et al. (2009) • Kumaraguru et al. (2008) • Jagatic et al. (2007) • Dodge et al. (2007) • Jackson, Ferguson, & Cobb (2005)	• Can directly and reliably observe the responses/behaviour. • High ecological validity. • Results are highly generalizable. • Fairly easy to stage. • Can target large populations. • Rich data can be collected using backend tools and scripts.	• It is difficult to get organizations willing to approve the staging of phishing attacks. • It is difficult to obtain research approvals and informed consent from participants. • Care has to be taken to ensure participants are not exposed to actual harm. • Technical expertise needed to deliver realistic/believable phishing attacks.

METHODOLOGY

This research used a naturalistic field study experiment to stage a phishing attack targeting a university population. This methodology is argued, with reasons summarized in Table 1, to be the most effective methodology when compared to the use of phishing knowledge IQ tests or lab experiments.

Research Setting

Previous researchers have found it very difficult to obtain cooperation to study information security threats in organizations (Bakhshi et al., 2009; Finn & Jakobsson, 2007; Huber et al., 2009; Vishwanath et al., 2011; Wang et al., 2012). This is a source of frustration for many information security researchers because many organizations either decline to have the study conducted altogether or restrict the publication of results (Kumaraguru et al., 2008). Some researchers opt not to conduct some elements of their study in order to obtain research approvals (Huber et al., 2009). In other cases, the research is prematurely terminated thereby negatively impacting data collection (Bakhshi et al., 2009).

There could be many reasons for this reluctance. Many organizations are wary of opening their doors for research due to the sensitivity of their systems and the confidential nature of their information and work practices (Burstein, 2008). They may not want their practices to be known to external parties, particularly competitors (they might lose intellectual property or competitive edge) or regulatory bodies (if they think their practices are deficient and may attract penalties). In addition, organizations are wary of negative publicity that may impact their bottom line due to loss of customers and revenue.

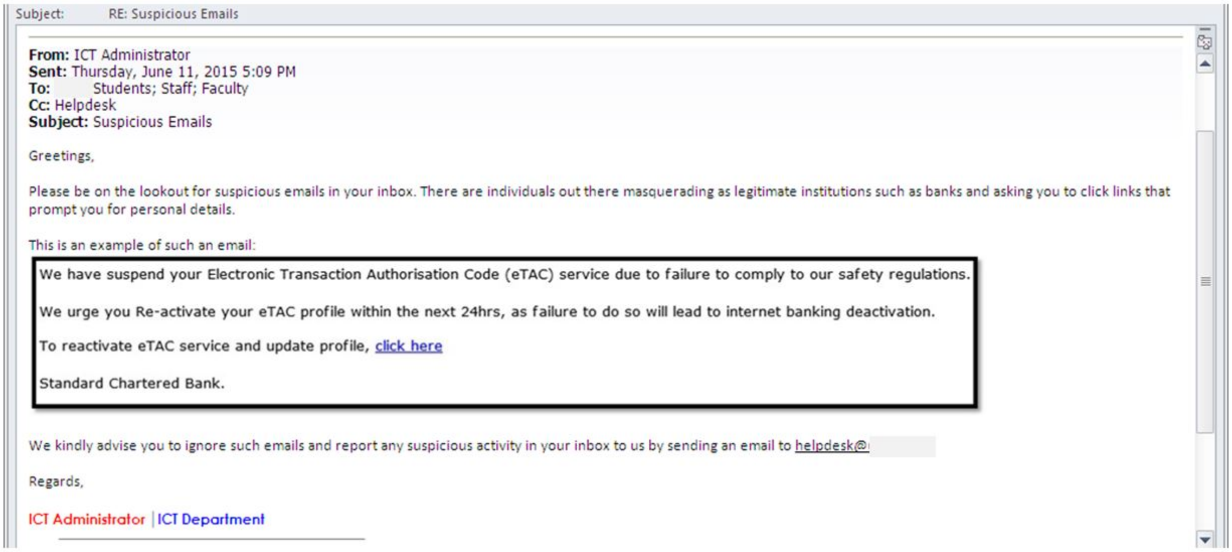
Therefore, a key criteria for selection of a research setting to study information security threats, such as phishing, is obtaining a willing organization that would give approval for conducting the research, the collection of sufficient data and publication of results (Bakhshi et al., 2009).

Getting a willing organization was an arduous task for this study. Five organizations consisting of 3 banks, 1 manufacturing company and 1 public utility company were contacted over a 14 month period to obtain approvals to conduct the research. All these institutions declined. The organization that was willing to allow this research to be conducted was a private university located in Nairobi, Kenya.

The selection of a university research site to study similar information security threats has been done in previous studies (Arachchilage & Love, 2013; Dodge et al., 2007; Finn & Jakobsson, 2007; Liang & Xue, 2010; Luo et al., 2013; Vishwanath et al., 2011; Wang et al., 2012; Workman, 2007, 2008a, 2008b). Universities are a suitable research site because they encourage research and the discovery of knowledge as long as the research is conducted ethically and does not harm the university community (Finn & Jakobsson, 2007).

Another key in determining the research setting was the selection of a naturalistic environment where the threat phenomenon was known to occur and could be observed without alerting study participants of the ongoing study. This required staging of attacks mimicking real-life threats and targeting study participants who were not aware of the ongoing research (Bakhshi et al., 2009; Finn & Jakobsson, 2007; Huber et al., 2009; Vishwanath et al., 2011). These staged attacks needed to be conducted in a way that made them as convincing and deceptive as real attacks.

The institution selected for this study had been a target of numerous social engineering attacks through phishing and wanted assistance in addressing the issue. Many of the attacks sought to obtain the confidential data, particularly passwords, through phishing emails as illustrated in Figure 1. Any identifying information in the figure has been greyed out to protect the identity of the institution.



Subject: RE: Suspicious Emails

From: ICT Administrator
Sent: Thursday, June 11, 2015 5:09 PM
To: [redacted] Students; Staff; Faculty
Cc: Helpdesk
Subject: Suspicious Emails

Greetings,

Please be on the lookout for suspicious emails in your inbox. There are individuals out there masquerading as legitimate institutions such as banks and asking you to click links that prompt you for personal details.

This is an example of such an email:

> We have suspend your Electronic Transaction Authorisation Code (eTAC) service due to failure to comply to our safety regulations.
>
> We urge you Re-activate your eTAC profile within the next 24hrs, as failure to do so will lead to internet banking deactivation.
>
> To reactivate eTAC service and update profile, click here
>
> Standard Chartered Bank.

We kindly advise you to ignore such emails and report any suspicious activity in your inbox to us by sending an email to helpdesk@[redacted]

Regards,

ICT Administrator | ICT Department

Figure 1: Sample phishing attack previously targeted at insiders

Other attacks sought to install malware on information systems through malicious attachments. The organization had been hit by numerous malware infections and ransomware attacks through this social engineering technique. The organization resonated with the proposed research and wanted assistance in assessing their exposure to the phishing threat.

Ethical Considerations

Research ethics relates to moral choices and decision making concerning research conduct (Greener, 2008). Various principles have to be upheld in the course of a research; these include: honesty, integrity, objectivity, respect for intellectual property, confidentiality and protection of research participants. Diener & Crandall (1978) highlight four main issues relating to research ethics: harm to participants, deception, invasion of privacy and lack of informed consent. Finn & Jakobsson (2007) point out that there are various ethical issues to be addressed when conducting information security research particularly when staging naturalistic experiments involving deception. This research took special care to address these ethical concerns.

Institutional approval to conduct research at the university was obtained from its research office and information communication and technology (ICT) department. This provided a site approval to conduct the research and collect data from the insiders. In addition, an Institutional Research Board (IRB) gained approval of the research proposal, data collection procedures and methodologies.

These various levels of approval were necessary in order to ensure that the study did not pose any actual harm to the participants or the institution. Two senior ICT administrators were attached to the research to review the phishing instruments to ensure that none of the technical components harmed the organization's information system or collected sensitive data from the insiders. The IRB approval signified that the research was found to meet the required ethical standards and was not going to be harmful to the participants or the organization.

Finn & Jakobsson (2007) point out that the deceptive nature of naturalistic studies makes it difficult to obtain informed consent from participants. This was also true for this study. However, the site approvals and oversight granted by the research office, ICT department and IRB protected the participants from adverse effects.

Diener & Crandall (1978) differentiate confidentiality and privacy and emphasize the need for research to fulfill these two key ethical considerations. Confidentiality is upheld in all stages of the research by making sure that study participants are anonymized, and no data is personally identifiable to them. Privacy regards the usage of the research data and this study ensures that the detailed raw data is not disclosed to other entities other than the researcher and the appointed academic supervisory teams. In addition, and published results are reported in collective terms where the organization or study participants are not identifiable. This study was bound by confidentiality and privacy requirements. Therefore, participant and institution data was anonymized and reported in collective terms.

Phishing Instruments

The development of the phishing instruments for this study was guided by the recommendations and lessons learnt from previous studies by Luo et al. (2013), Arachchilage & Love (2013), Vishwanath et al. (2011) and Bakhshi et al. (2009).

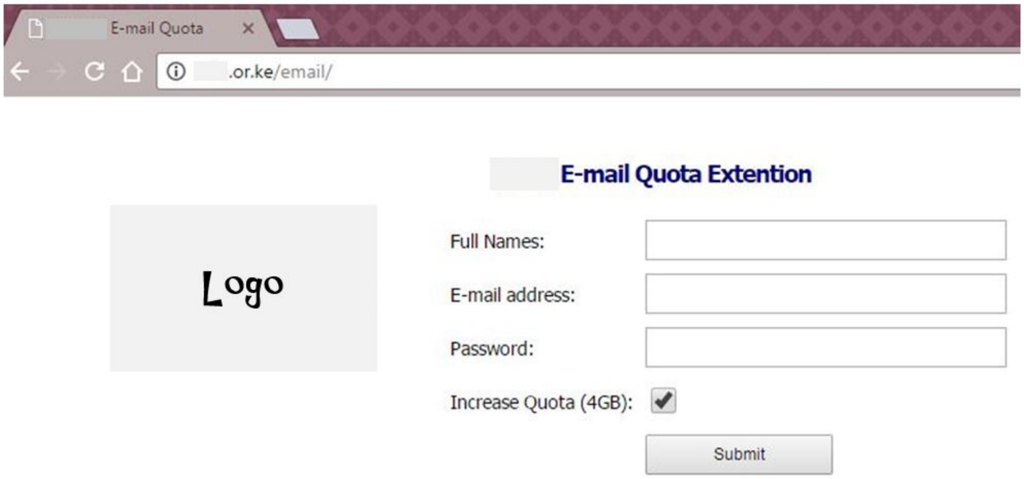
First, typical samples of phishing attacks launched against the insiders in the organization were studied. The ICT administrators attached to the research provided 12 samples of recent phishing attacks that had been targeted at the organization's insiders. Characteristics that made the phishing attacks successful were identified in collaboration with the ICT administrators. The attacks that closely imitated the organization's communication techniques and the look and feel were seen to be most deceptive. Therefore, the phishing instruments were designed to closely conform to the layout, fonts, look and feel used within the organization.

Secondly, a domain that imitates the organization's domain was selected. Instead of using the registered domain ending with "ac.ke," the researcher registered a domain that ended with "or.ke." The email address "helpdesk@universityX.or.ke" was used and the website was hosted on "universityX.or.ke." This ensured that the email and website address used to conduct the attack would closely imitate the organization’s legitimate addresses while allowing for knowledgeable insiders to identify the attack by picking up an inconsistency in the addressing. This strategy is advocated by Luo et al. (2013).

The next step in the process involved the selection of a pretext scenario that would be perceived as a natural event. The scenario would then guide the development of content for the phishing email and message. The guidelines by Luo et al. (2013) and Vishwanath et al. (2011) were used to guide the design of the pretext scenario. A topic that was current and relevant to the organization was selected. The organization had a limited capacity email server and consequently users were only allowed 2GB of email space. This meant that users regularly received ‘mailbox full’ notifications indicating they had exhausted their allocated quota. The pretext scenario took advantage of this and advertised an opportunity for the users to increase their allocated email quota. Time pressure was also put on the users to respond urgently in order to prevent discontinuation of service similar to the Luo et al. (2013) study.

A data collection website developed in HTML5, CSS and PHP with a MySQL database was hosted on the registered domain and tested to ensure it ran without errors. In addition, the ICT administrators attached to the study reviewed the code and backend database to ensure that no malware was delivered and no sensitive or confidential data was collected and stored. This protected the insiders from actual harm as was required by directives from the research office and Institutional Review Board.

Figure 2 depicts the login page of the website. Appendix I provides the source code and Appendix II provides the Cascading Style Sheets (CSS) for the page. Any identifying information has been removed from the content to protect the identity of the institution.



E-mail Quota Extention

Full Names:

E-mail address:

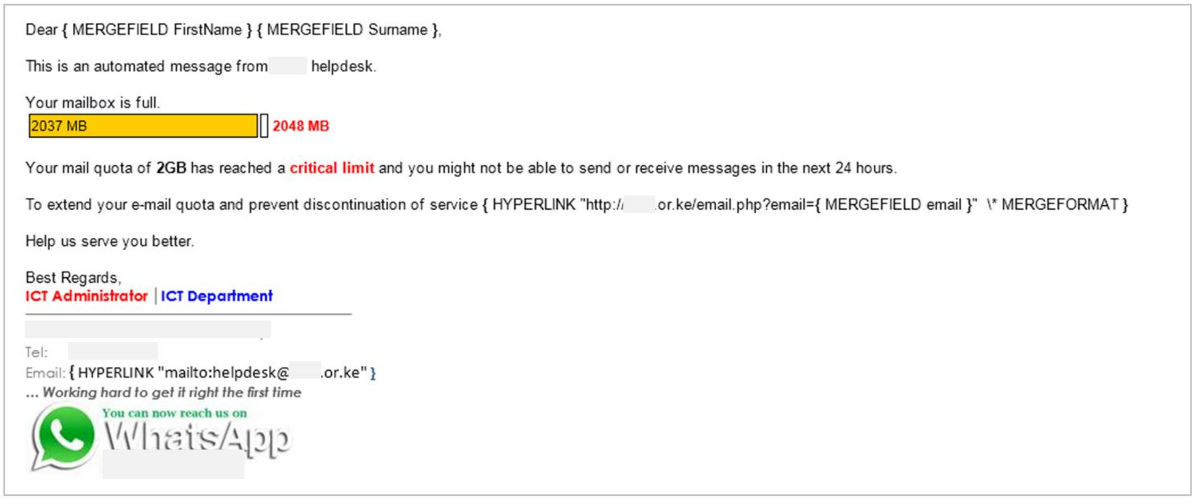
Password:

Increase Quota (4GB): ☑

Figure 2: Phishing Website Login

Next, targeted phishing emails were sent to selected insiders. The emails were staged as spear phishing emails using the first name and surname to personalize the message. The message seemed to have been sent from the institution’s helpdesk by an ICT administrator. This imitated the means of communication commonly used by the institution when sending IT related information to the users.

The email had the ‘look’ and ‘feel’ of the usual email messages from ICT administrators. It was carefully composed not to have spelling mistakes or sloppy content so that recipients do not superficially dismiss it. The variable fields in the email were filled in using mail merge. These fields were: first name, last name and email address. Figure 3 shows the mail merge template setup using the mail merge feature on Microsoft Office Word 2013. Any identifying information has been greyed out to protect the identity of the institution.



Dear { MERGEFIELD FirstName } { MERGEFIELD Surname },

This is an automated message from [redacted] helpdesk.

Your mailbox is full.
2037 MB 2048 MB

Your mail quota of **2GB** has reached a **critical limit** and you might not be able to send or receive messages in the next 24 hours.

To extend your e-mail quota and prevent discontinuation of service { HYPERLINK "http://[redacted].or.ke/email.php?email={ MERGEFIELD email }" * MERGEFORMAT }

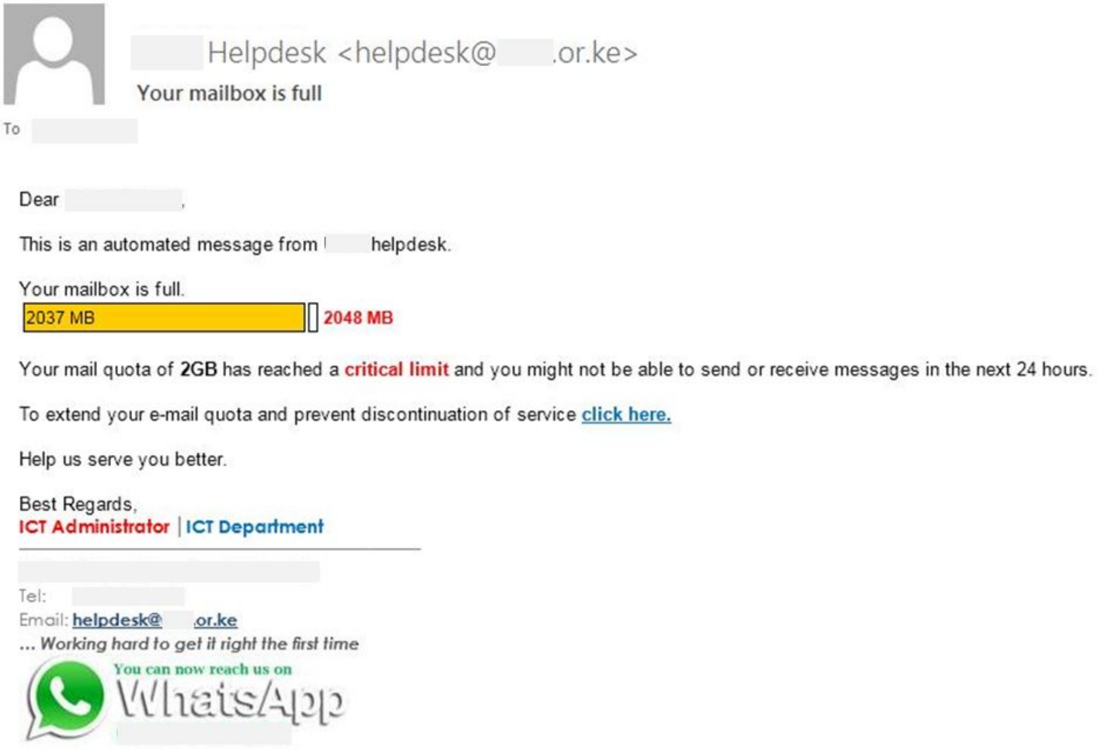
Help us serve you better.

Best Regards,
ICT Administrator | **ICT Department**

[redacted]
Tel: [redacted]
Email: { HYPERLINK "mailto:helpdesk@[redacted].or.ke" }
... Working hard to get it right the first time
You can now reach us on WhatsApp
[redacted]

Figure 3: Phishing Mail Merge Template

The administration of emails was automated using mail merge working together with Microsoft Outlook 2013. Figure 4 shows the resulting phishing email that was sent to a sample of targeted insiders. Please note that identifying information has been greyed out to protect the identity of the institution.



[redacted] Helpdesk <helpdesk@[redacted].or.ke>
Your mailbox is full
To [redacted]

Dear [redacted],

This is an automated message from [redacted] helpdesk.

Your mailbox is full.
2037 MB 2048 MB

Your mail quota of **2GB** has reached a **critical limit** and you might not be able to send or receive messages in the next 24 hours.

To extend your e-mail quota and prevent discontinuation of service click here.

Help us serve you better.

Best Regards,
ICT Administrator | **ICT Department**

[redacted]
Tel: [redacted]
Email: helpdesk@[redacted].or.ke
... Working hard to get it right the first time
You can now reach us on WhatsApp
[redacted]

Figure 4: Phishing Email

These phishing instruments collected various data items for study. The phishing email tracked when the email was successfully delivered and opened. In addition, the phishing email had a hyperlink in which the words “click here” were highlighted in blue and underlined. This hyperlink did two things. First, it directed the person to the phishing website by opening their default browser and loading the phishing website’s address. Secondly, it passed on a unique identifier as a pre-filled parameter to the landing page. This means it was possible to distinctively track all the people who visited the website.

The phishing website ran active scripts that recorded a timestamp of when the page was loaded, the identifier registered from the forwarding email and various parameters about the system accessing the page including the IP address, browser and Operating System. The source code of the background script is provided in Appendix III. This means that even if the user did not interact further with the website, just loading it gave a lot of valuable information.

The other way data was collected was when a person filled in the form on the website. This involved submitting the following details: full name, email address and password. The email address was already pre-filled if the person clicked the hyperlink from the phishing email. This communicated some level of sophistication to users that was designed to make the website more trustworthy. When a person filled in the form and clicked the submit button their password was neither captured nor transmitted as a design requirement. This prevented the capturing of confidential information and protected the institution from actual harm. The webpage also had error validation to ensure that the submit functionality did not work if the required form fields were blank.

Sampling

In the context of this study, the effective population was all the insiders who had active email accounts on the university’s system. These were all the potential targets of any phishing attack directed at the university using its domain. The domain account management system was queried by its system administrator to provide the exact number of insiders at the time of the study.

Strata	Number
Students	7,729
Staff	312
Adjunct Faculty	158
Full-time Faculty	141
Management	13
Interns	9
Mailing List Users	7
Unknown	36
Total Insiders	**8,405**

Table 2: Sampling Frame

The university campus had a total of 8,405 insiders active on its information system. Of these, 7,729 were students, 312 were staff members, 158 were adjunct faculty, 141 were full-time faculty, 13 were management, 9 were interns, 7 accounts were mailing list accounts and 36 could not be classified in any of these categories due to insufficient metadata. Table 2 illustrates this sampling frame.

This study employed a probability sampling technique to allow the results to be generalizable to the population. Bhattacherjee (2012) explains that in probability sampling, each entity in the population has a non-zero chance of being selected in the sample. In addition, random selection techniques are employed in the sampling process. This ensures that sample statistics are unbiased estimates of what is in the population.

The specific technique selected was proportional stratified random sampling. The process as outlined by Bhattacherjee (2012) involves dividing the sampling frame into non-overlapping groups called strata. Thereafter a simple random sample is drawn from each stratum in what is called multi-stage random sampling. This ensures that the strata with few members is not oversampled and the resulting sample has similar ratios for the different strata.

The determination of sample size used the Cochran (1977) formula. It targeted a 95% confidence level and a very low margin of error at 1%. The proportion of sampling in the population was set at 50% to give maximum variability. This resulted in a sample size of 4,483 being extracted from the population of 8,405 insiders. To prevent under-sampling or over-sampling per strata, proportional stratified random sampling was done to determine the actual composition of the sample per strata. The numbers per strata selected for the sample are represented in Table 3.

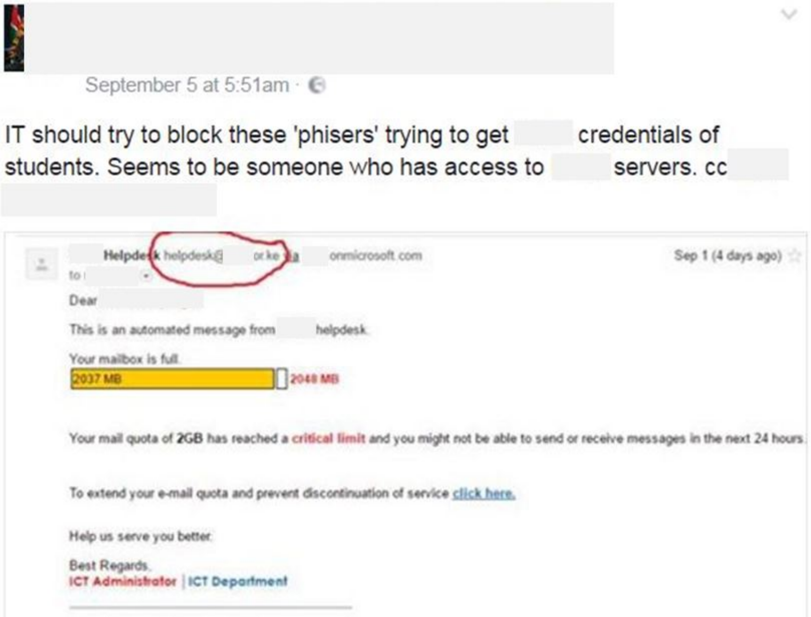
Strata	Number	Proportion	Size in Sample
Students	7,729	91.96%	4,122
Staff	312	3.71%	166
Adjunct Faculty	158	1.88%	84
Full-time Faculty	141	1.68%	75
Management	13	0.15%	6
Interns	9	0.11%	4
Mailing List Users	7	0.08%	7
Unknown	36	0.43%	19
Total	**8,405**	**100%**	**4483**

Table 3: Sample Size

The size in sample for each stratum was then chosen using simple random sampling with the aid of a random number generator. To do this, the dataset associated with the 8,405 users were loaded onto a Microsoft Excel 2013 workbook. Each row of the workbook was associated with one user. The entries were grouped sequentially according to the strata outlined in Table 3. Next, a new column was added on the workbook to contain the random number. The random number was generated using the RAND() function entered as a formula =RAND() for every cell in the column. This ensured that each user entry was assigned a random number. After the random numbers were assigned, the entries were sorted in ascending order while still maintaining the strata groupings. Finally, the required size in sample, say 'n_s', was selected by choosing the first n_s entries in each stratum. These entries were transferred to a new workbook representing the selected sample dataset of 4,483 users.

RESULTS AND ANALYSIS

The phishing experiment ran for 40 days. It had to be stopped because a prominent social media activist and blogger, who was also a student at the university, called for the phishing to be investigated and stopped. His comment was posted on the university's social media and within a few hours had reached many people within the university. The social media post is illustrated in Figure 5. Any identifying information has been greyed out to protect the identity of the institution.



[redacted]

September 5 at 5:51am ·

IT should try to block these 'phisers' trying to get [redacted] credentials of students. Seems to be someone who has access to [redacted] servers. cc [redacted]

[redacted] Helpdesk helpdesk@[redacted] ac.ke via [redacted] onmicrosoft.com Sep 1 (4 days ago)

to [redacted]

Dear [redacted]

This is an automated message from [redacted] helpdesk.

Your mailbox is full.

2037 MB 2048 MB

Your mail quota of **2GB** has reached a **critical limit** and you might not be able to send or receive messages in the next 24 hours.

To extend your e-mail quota and prevent discontinuation of service click here.

Help us serve you better.

Best Regards,
ICT Administrator | ICT Department

Figure 5: Phishing alert sent on social media

This prompted the administration at the university to call off the exercise due to the alarm raised. The ICT director, who had been involved in the research approvals and was aware of its progress, instructed his team to send out alerts to the entire university community informing them of the nature of the research and allaying any concerns of an actual threat. This demonstrates the power informed and vocal insiders have in identifying threats and alerting their communities to curtail targeted attacks.

By this time, all the 4,483 insiders who were sampled from the university community had already been sent phishing emails through their official university accounts. The email system returned delivery failures for 138 of the emails indicating that there was a problem with these email accounts. This meant that 4,345 phishing emails were delivered to the insiders' official email accounts. Statistics on interaction with the phishing email were low. There was no response or interaction with the phishing email by 4,104 of the targeted sample.

Category	Number	Percentage
Sample size targeted with phishing email	4,483	**100%**
E-mail delivery failures	138	**3.08%**
Did not read/interact with phishing email	4,104	**91.54%**
Insiders that participated	**241**	**5.37%**

Table 4: Response Rate Statistics

The number of insiders who participated in the experiment were 241. This was 5.37% of the total number sampled. These are the people who received the phishing emails and opened them. Read receipts were setup in Microsoft Outlook to give this indication. Data collected on the backend database indicated a total of 98 clicks on the phishing hyperlink. These clicks were associated with 74 unique insiders since some clicked the phishing hyperlink multiple times, as indicated by repeated entries in the backend database. In addition, the form on the phishing website was filled in 72 times with 65 form-fills being unique and the others being repeated entries. This shows 87.84% of the insiders who were susceptible to phishing emails went ahead to disclose passwords that would enable an attacker gain access to the organization's systems.

The response rate per strata is provided in Table 5. Results shows that interns (25%), staff (22.89%), full-time faculty (17.33%) and mailing list users (14.29%) had higher response rates in proportion to the numbers targeted per strata. Students had a very low percentage (0.49%) of successfully phished per strata despite having the highest number in sample.

Strata	Size in Sample	Successfully Phished	Proportion
Students	4,122	20	0.49%
Staff	166	38	22.89%
Adjunct Faculty	84	1	1.19%
Full-time Faculty	75	13	17.33%
Management	6	0	0%
Interns	4	1	25%
Mailing List Users	7	1	14.29%
Unknown	19	0	0%
Total	**4483**	**74**	**100%**

Table 5: Response Rate per Strata

DISCUSSION AND RECOMMENDATIONS

The study presents interesting findings and valuable lessons. These lessons are hereafter synthesized into guidelines for practice. The first guideline advocates for the use of naturalistic field experiments when assessing how susceptible insiders are to phishing attacks. This is because naturalistic studies allow a direct observation of actual behaviour and this provides a more reliable assessment of phishing susceptibility than self-reported measures. The insiders being studied are not alerted about the assessment and are expected to act as they normally would in their real-world settings. This provides a high ecological validity of results and makes them generalizable to the population. In addition, rich data informing researchers on user behaviours can be collected from large populations with relative ease.

The second guideline emphasizes the need to obtain research approvals from the organization where the study is to be conducted. Getting permission to conduct information security research is often an arduous task, as was with this study. This however does not preclude the need for research approvals. It is important to get site approvals from the necessary representatives of the institution where the research will be conducted. It is important to obtain an ethical review approval from an IRB to ensure that the research protocol protects participants from actual harm. In this study, approvals were obtained from the university's research office, ICT department and IRB. These layers of review protected the institution and its insiders from adverse effects during the staged phishing attack.

The third guideline relates to the development and setup of phishing instruments. In this study, phishing was conducted using targeted spear phishing emails and also by setting up a phishing website. Before any phishing instrument was developed care was taken to design them to be convincing. Previous phishing attacks targeted at the institution were studied and the characteristics of regular communication were noted. In addition, a pretext scenario that was relevant to the current affairs at the organization was chosen. These considerations during design ensure that the staged attack is not easily dismissed without eliciting interaction from those targeted. In addition, a phishing domain that was deceptively similar to the organization's operational domain was registered. Organizations should closely monitor domains that are very similar to their operational domains. These could be deceptive variants of their operational domain but also those ending with different suffixes such as .com, .org or even country suffices such as .or.ke, as was used in this study. The information technology or security teams at the organization should probably go a step further to buy such domains instead of leaving them available for outsiders to acquire. It is not a very expensive venture since registering a domain could cost as low as 10 dollars per year, as was the case in this study. This study also setup a phishing email address to imitate the ICT helpdesk correspondence to deliver spear phishing emails. The phishing emails and website used in this study were designed with active scripts that did a lot of background work. This highlights a very important point. Phishing is not considered successful only when a person fills in sensitive information on a web form. Attackers collect valuable information right from the time a person opens their emails or loads their websites. Current phishing scams are very sophisticated. Emails and webpages contain a lot of active scripts that harvest information from user systems and even install malware without visibly alerting users. A key contribution of this research involves presenting the actual code that was used to implement the phishing webpages and login forms, the active background scripts that harvested system details and also the mail merge templates that were used to deliver phishing emails.

The fourth guideline relates to population sampling. Care should be taken to ensure a representative sample is drawn from the population before the staged phishing attack is delivered. This study advocates for a probability sampling technique because it allows the results of the assessment to be generalizable to the wider population under investigation. The study outlines the sampling process in detail. It starts by extracting a dataset of user accounts from the information system and arranging these accounts according to the different functional divisions in the organization. This ensures that no division or functional group is over-sampled or under-sampled. Thereafter, actual phishing targets are selected randomly and in proportion to the strata representation in the population. This rigorous sampling process ensures the sample used in the assessment is a good representation of the population and this in turn allows the results and lessons learnt to be generalized to the wider organization without bias.

The fifth guideline relates to the actual execution of the staged phishing attack. In this study, the staged attack ran for 40 days. It is important to allow a similar amount of time for the exercise to run to account for differences in email responsiveness among the population. Some participants may respond immediately while others may defer their email processing for days or even weeks. The days allocated to the exercise should also factor times where participants are expected to be on holiday or may have limited access to their accounts. During the execution it would be important to look out for any insiders who alert the community about the attack and to examine the reaction of the organization to this alert. These vocal insiders could affect the event of an actual attack and could present a very important countermeasure that organizations should focus on when addressing attacks. In this study, a prominent blogger was able to raise an alarm and rally action through social media. Within a few hours, an alert of the ongoing threat had been circulated throughout the entire institution. Study protocols should collect data and assess the effectiveness of such countermeasures in curtailing attacks. Organizations should invest in channels through which users can quickly report suspected attacks and through which information can be shared with the wider population to frustrate the efforts of attackers. They should turn each user on their system into an intrusion detection agent with the skill and capability to detect threats and to sound an alarm for action.

The sixth guideline relates to the actual results of the study. This study suffered a significant attrition with regards to the actual users who engaged with the phishing emails. Only 5.37% of the 4,483 sampled participants engaged with the phishing email. This is comparable to the study by Mohebzada, El Zarka, BHojani, & Darwish (2012) where over 10,000 phishing emails were sent to faculty, staff, students and alumni of a university. Two types of phishing emails were sent. The first phishing email had an 8.74% success rate and the second 2.05%. Low phishing rates could be an indication that the participants identified the phishing scam and chose not to engage with the phishing email or website. Although the phishing rates were low, it only takes a few users to compromise an information system. Once an attacker is successful with some systems, these can then be used as a pivot point to work into the rest of the organization (Ali, 2015).

Some lessons can be learnt from our study to help future studies increase the number of users who interact with the phishing instruments. Firstly, it would be important to confirm that the emails that will be targeted are operational and that no delivery failures will be experienced. Secondly, it is important to confirm whether users regularly engage with their emails. Discussions with the ICT staff attached to the study revealed that it could be that few people used their official university accounts for correspondence. Students (who were the largest number in the sample) had an option of using alternative email addresses to receive communication from the university. This meant that they had no imperative to use their official email accounts. Instead they preferred to use private email accounts mainly from Google, Hotmail or Yahoo. If official institution email addresses are not used regularly, then personal emails registered for official communication should also be included in the sample. Thirdly, increasing the study period would also give users a longer time to review their emails and thereby possibly increase their participation. Fourthly the assessment could also target other channels to deliver the phishing attack, for example, using organizational social media accounts and telephone chat groups.

CONCLUSION

Phishing is still a very prevalent form of social engineering attack leveraged against organizations today. Recent reports have shown that it is a common method of compromising organizations and spreading Advanced Persistent Threats (APTs). It is important that organizations take steps to assess their level of risk and exposure to this attack. This study presents a way in which organizations can use a naturalistic study to objectively assess their exposure to phishing threats. Organizations can use such naturalistic experiments to regularly determine the extent to which their users can succumb to phishing attacks. The data collection instruments used in the naturalistic field study are not difficult to assemble. This study makes an important contribution by outlining the actual tools used to stage the phishing attack in detail. Such assessments can be run on a routine basis to provide a security baseline metric from which to compare from time to time. The results of the assessments can be very useful in designing countermeasures, one of which is discussed in this study. Insiders can be equipped to detect attacks and channels to alert the wider community can be provided to them. This would inevitably provide an essential component of strengthening the overall information security of an organization, particularly from a people-perspective, which organizations often leave unaddressed.

ACKNOWLEDGMENTS

This work was funded through the German Academic Exchange Service (DAAD) In-Country/ In-Region Scholarships for Postgraduates.

REFERENCES

Ali, A. (2015). Social Engineering: Phishing latest and future techniques. Retrieved March 10, 2016 from https://www.researchgate.net/publication/274194484

Anandpara, V., Dingman, A., Jakobsson, M., Liu, D., & Roinestad, H. (2007). Phishing IQ Tests Measure Fear, Not Ability (pp. 362–366). Presented at the International Conference on Financial Cryptography and Data Security, Trinidad and Tobago: Springer Berlin Heidelberg.

APWG, A.-P. W. G. (2016). Phishing Activity Trends Report: 1st Quarter 2016. Anti-Phishing Working Group. Retrieved March 9, 2016 from https://docs.apwg.org/reports/apwg_trends_report_q1_2016.pdf

APWG, A.-P. W. G. (2017). Phishing Activity Trends Report: 4th Quarter 2016. Anti-Phishing Working Group. Retrieved December 12, 2017 from https://docs.apwg.org/reports/apwg_trends_report_q4_2016.pdf

APWG, A.-P. W. G. (2018). Phishing Activity Trends Report: 3rd Quarter 2017. Anti-Phishing Working Group. Retrieved April 30, 2018 from http://docs.apwg.org/reports/apwg_trends_report_q3_2017.pdf

Arachchilage, N. A. G., & Love, S. (2013). A Game Design Framework for Avoiding Phishing Attacks. *Computers in Human Behavior*, *29*, 706–714.

Bakhshi, T., Papadaki, M., & Furnell, S. (2009). Social Engineering: Assessing Vulnerabilities in Practice. *Information Management & Computer Security*, *17*(1), 53–63. https://doi.org/10.1108/09685220910944768

Barth, B. (2016). Don’t be like “Mike”: Authorities arrest mastermind of $60M online scam operation. SC Magazine.

BBC News. (2016). Online fraud: Top Nigerian scammer arrested. BBC News. Retrieved October 3, 2016 from http://www.bbc.com/news/world-africa-36939751

Bhattacherjee, A. (2012). *Social Science Research: Principles, Methods, and Practices* (2nd ed.). University of South Florida Scholar Commons.

Brewer, M. B., & Crano, W. D. (2014). Research Design and Issues of Validity. In H. T. Reis & C. M. Judd (Eds.), *Handbook of Research Methods in Social and Personality Psychology* (2nd ed., pp. 3–16). New York: Cambridge University Press.

Burstein, A. J. (2008). *Toward a Culture of Cybersecurity Research* (UC Berkeley Public Law Research Paper No. 1113014). Retrieved from http://dx.doi.org/10.2139/ssrn.1113014

CERT, I. T. T. (2013). Unintentional Insider Threats: A Foundational Study. Software Engineering Institute, Carnegie Mellon University. Retrieved October 17, 2013 from http://resources.sei.cmu.edu/asset_files/TechnicalNote/2013_004_001_58748.pdf

Cimpanu, C. (2016). Anonymous Hackers Leak 1 TB of Documents from Kenya’s Ministry of Foreign Affairs. Retrieved September 12, 2016, from http://news.softpedia.com/news/anonymous-hackers-leak-1tb-of-documents-from-kenya-s-ministry-of-foreign-affairs-503518.shtml

Cochran, W. G. (1977). *Sampling Techniques* (3rd ed.). New York: Wiley.

Cyveillance. (2015). The Cost of Phishing: Understanding the True Cost Dynamics Behind Phishing Attacks.

Dhamija, R., Tygar, J. D., & Hearst, M. (2006). Why Phishing Works. In *Proceedings of the SIGCHI conference on Human Factors in computing systems* (pp. 581–590). ACM.

Diener, E., & Crandall, R. (1978). *Ethics in Social and Behavioral Research*. University of Chicago Press.

Dodge, R. C., Carver, C., & Ferguson, A. J. (2007). Phishing for User Security Awareness. *Computers & Security*, *26*, 73–80.

Downs, J. S., Holbrook, M., & Cranor, L. F. (2006). Decision Strategies and Susceptibility to Phishing. In *Proceedings of the second symposium on Usable privacy and security* (pp. 79–90). ACM.

Downs, J. S., Holbrook, M., & Cranor, L. F. (2007). Behavioral Response to Phishing Risk. In *Proceedings of the anti-phishing working groups 2nd annual eCrime researchers summit* (pp. 37–44). ACM.

Egelman, S., Cranor, L. F., & Hong, J. (2008). You’ve been warned: an empirical study of the effectiveness of web browser phishing warnings. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM.

Ferguson, A. J. (2005). Fostering e-mail security awareness: The West Point Carronade. *Educase Quarterly*, *28*(1), 54–57.

Finn, P., & Jakobsson, M. (2007). Designing and Conducting Phishing Experiments. *IEEE Technology and Society Magazine, Special Issue on Usability and Security*, *26*(1), 46–58.

Fire Eye. (2015). APT30 And The Mechanics Of A Long-Running Cyber Espionage Operation. Retrieved August 16, 2017 from https://www.fireeye.com/current-threats/threat-intelligence-reports.html

Fire Eye. (2017). APT28: At The Center Of The Storm. Retrieved August 16, 2017 from https://www.fireeye.com/current-threats/threat-intelligence-reports.html

Greener, S. (2008). *Business Research Methods*. BookBoon.

Hernandez, E., Regalado, D., & Villeneuve, N. (2015). An Insider Look Into the World of Nigenrian Scammers. Fire Eye. Retrieved August 16, 2017 from https://www.fireeye.com/current-threats/threat-intelligence-reports.html

Huber, M., Kowalski, S., Nohlberg, M., & Tjoa, S. (2009). Towards Automating Social Engineering Using Social Networking Sites (Vol. 3, pp. 117–124). Presented at the Computational Science and Engineering, IEEE.

Jackson, J. W., Ferguson, A. J., & Cobb, M. J. (2005). Building a University-wide Automated Information Assurance Awareness Exercise. (p. T2E7-11). Presented at the 35th ASEE/IEEE Frontiers in Education Conference, Indianapolis, IN, USA: IEEE.

Jagatic, T., Johnson, N., Jakobsson, M., & Menczer, F. (2007). Social Phishing. *Communications of the ACM*, *50*(10), 94–100.

Jakobsson, M. (2007). The Human Factor in Phishing. *Privacy & Security of Consumer Information*, *7*(1), 1–19.

Jakobsson, M., Tsow, A., Shah, A., Blevis, E., & Lim, Y.-K. (2007). What Instills Trust? A Qualitative Study of Phishing. In *Financial Cryptography and Data Security* (pp. 356–361). Springer Berlin Heidelberg.

James, L. (2005). *Phishing Exposed*. Syngress.

Kumaraguru, P., Cranshaw, J., Acquisti, A., Cranor, L., Hong, J., Blair, M. A., & Pham, T. (2009). School of Phish: A Real-Word Evaluation of Anti-Phishing Training. In *Proceedings of the 5th Symposium on Usable Privacy and Security (SOUPS)*. Mountain View, CA, USA: ACM.

Kumaraguru, P., Rhee, Y., Acquisti, A., Cranor, L. F., Hong, J., & Nunge, E. (2007). Protecting People from Phishing: The Design and Evaluation of an Embedded Training Email System. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 905–914). ACM.

Kumaraguru, P., Rhee, Y., Sheng, S., Hasan, S., Acquisti, A., Cranor, L. F., & Hong, J. (2007). Getting Users to Pay Attention to Anti-Phishing Education: Evaluation of Retention and Transfer. In *Proceedings of the anti-phishing working groups 2nd annual eCrime researchers summit* (pp. 70–81). ACM.

Kumaraguru, P., Sheng, S., Acquisti, A., Cranor, L. F., & Hong, J. (2007). Teaching Johnny Not to Fall for Phish. *CM Transactions on Internet Technology (TOIT)*, *10*(2), 7.

Kumaraguru, P., Sheng, S., Acquisti, A., Cranor, L., & Hong, J. (2008). Lessons From a Real World Evaluation of Anti-Phishing Training. Presented at the eCrime Researcher's Summit, Anti-Phishing Working Group (APWG).

Liang, H., & Xue, Y. (2010). Understanding Security Behaviors in Personal Computer Usage: A Threat Avoidance Perspective. *Journal of the Association for Information Systems (JAIS)*, *11*(7), 394–413.

Luo, X. (Robert), Brody, R., Seazzu, A., & Burd, S. (2011). Social Engineering: The Neglected Human Factor for Information Security Management. *Information Resources Management Journal (IRMJ)*, *24*(3), 1–8.

Luo, X. (Robert), Zhang, W., Burd, S., & Seazzu, A. (2013). Investigating Phishing Victimization with the Heuristic-Systematic Model: A Theoretical Framework and an Exploration. *Computers & Security*, *38*, 28–38.

Mandiant. (2004). APT1: Exposing One of China's Cyber Espionage Units - FireEye. Mandiant. Retrieved October 6, 2016 from https://www.fireeye.com/content/dam/fireeye-www/services/pdfs/mandiant-apt1-report.pdf

Mandiant. (2010). M-Trends: The Advanced Persistent Threat. Mandiant. Retrieved October 6, 2016 from https://www.fireeye.com/blog/threat-research/2010/01/m-trends-advanced-persistent-threat-malware.html

Mohebzada, J. G., El Zarka, A., BHojani, A. H., & Darwish, A. (2012). Phishing in a University Community: Two Large Scale Phishing Experiments (pp. 249–254). Presented at the International Conference on Innovations in Information Technology (IIT), IEEE.

Obulutsa, G. (2016). Hackers leak stolen Kenyan foreign ministry documents. Retrieved September 12, 2016 from http://www.reuters.com/article/us-cyber-kenya-idUSKCN0XP2K5

Parsons, M. H. (1974). What happened at Hawthorne? *Science*, *183*(4128), 922–932.

PhishTank. (2016). PhishTank Stats. Retrieved October 14, 2016 from https://www.phishtank.com/stats.php

Tsow, A., & Jakobsson, M. (2007). Deceit and Deception: A Large User Study of Phishing. Indiana University.

Verizon. (2015). 2015 Data Breach Investigations Report (DBIR). Retrieved July 16, 2015 from http://news.verizonenterprise.com/2015/04/2015-verizon-dbir-report-security/

Verizon. (2016). 2016 Data Breach Investigations Report (DBIR). Retrieved July 16, 2016 from http://www.verizonenterprise.com/resources/reports/rp_DBIR_2016_Report_en_xg.pdf

Verizon. (2017, April). 2017 Data Breach Investigations Report (DBIR). Retrieved May 16, 2017 from http://www.verizonenterprise.com/verizon-insights-lab/dbir/2017/

Vishwanath, A., Herath, T., Chen, R., Wang, J., & Rao, R. H. (2011). Why do people get phished? Testing individual differences in phishing vulnerability within an integrated, information processing model. *Decision Support Systems*, *51*(3), 576–586.

Wang, J., Herath, T., Chen, R., Vishwanath, A., & Rao, R. H. (2012). Phishing Susceptibility: An Investigation Into the Processing of a Targeted Spear Phishing Email. *IEEE Transactions on Professional Communication*, *55*(4), 345–362.

Waqas. (2016). Anonymous Leaks 1TB of Data from Kenya's Ministry of Foreign Affairs. Retrieved September 12, 2016 from https://www.hackread.com/anonymous-hacks-kenya-ministry-foreign-affairs/

Workman, M. (2007). Gaining Access with Social Engineering: An Empirical Study of the Threat. *Information Systems Security*, *16*(6), 315–331.

Workman, M. (2008a). A Test of Interventions for Security Threats from Social Engineering. *Information Management & Computer Security*, *16*(5), 463–483.

Workman, M. (2008b). Wisecrackers: A Theory-Grounded Investigation of Phishing and Pretext Social Engineering Threats to Information Security. *Journal of the American Society for Information Science and Technology*, *59*(4), 662–674.

APPENDIX I: INDEX HTML PAGE SOURCE CODE

```
<?php
        session_start();

        // Initialize variables
        $username="";
        $email="";
        $passwordErr="";
        $nameErr ="";
        $emailErr ="";
        $passwordErr ="";
        $isValidUsername=0;
        $isValidEmail = 0;
        $isValidPassword = 0;

        function test_input($data) {
                $data = trim($data);
                $data = stripslashes($data);
                $data = htmlspecialchars($data);
                return $data;
        }

        if($_SERVER["REQUEST_METHOD"] == "POST") {
                // Form submitted
```

```
//-------------------------Form Validation Start---------------------//
if (empty($_POST["username"])) {
        $nameErr = "Name is required";
        $isValidUsername = 0;
} else {
        $username = test_input($_POST["username"]);
        if (!preg_match("/^[a-zA-Z ]*$/",$username)) {
                $nameErr = "Only letters and white space allowed";
                $isValidUsername = 0;
        }
        else {
                $isValidUsername = 1;
        }
}

if (empty($_POST["email"])) {
        $emailErr = "E-mail is required e.g. username@uni.ac.ke";
        $isValidEmail = 0;
} else {
        $email = test_input($_POST["email"]);
        $regex = '/^[_a-z0-9-]+(\.[_a-z0-9-]+)*@[a-z0-9-]+(\.[a-z0-9-]+)*(\.[a-z]{2,4})$/';
        if (!preg_match($regex, $email)) {
                $emailErr = "$email is not a valid email address";
                $isValidEmail = 0;
        }
        else {
                $isValidEmail = 1;
        }
}

if (empty($_POST["password"])) {
        $passwordErr = "Password is required";
        $isValidPassword = 0;
} else {
        $password = md5($_POST["password"]);
        $isValidPassword = 1;
}

//-------------------------Form Validation End-----------------------//

//-------------------------Database Connection Start-----------------------//
if ($isValidUsername && $isValidEmail && $isValidPassword){
        //Set up connection to database
        define('DB_SERVER', 'SERVER_NAME');
        define('DB_USERNAME', 'USER_NAME');
        define('DB_PASSWORD', 'PASSWORD');
        define('DB_DATABASE', 'DB_NAME');
        $db = mysqli_connect(DB_SERVER,DB_USERNAME,DB_PASSWORD,DB_DATABASE);

        if (!$db) {
                die("Connection failed: " . mysqli_connect_error());
        }

        //mysqli_real_escape_string used to prevent SQLi
        $username = mysqli_real_escape_string($db,$username);
        $email = mysqli_real_escape_string($db,$email);

        //No password stored to protect users
        $sql = "INSERT INTO responses (`names`, `email`) VALUES ('$username','$email')";

        if (mysqli_query($db,$sql)){
                echo "Your email quota has been increased to 4GB";
```

```
                                echo '<script
                                type="text/javascript">window.location.href="http://www.UNI.ac.ke";</script>';
                                }
                                else {
                                echo "Error: " . $sql . "<br>" . mysqli_error($db);
                                }

                        mysqli_close($db);
                }
                //-------------------------Database Connection End---------------------//

        }

?>

//-------------------------HTML5 Index Page Start---------------------//

<!doctype html>
<html lang="en">

<head>
        <meta charset="utf-8">
        <title>E-mail Quota</title>
        <link rel="stylesheet" type="text/css" href="stylesheet.css">
</head>

<body topmargin='0' bottommargin='0' leftmargin='0' rightmargin='0' marginwidth='0' marginheight='0'
        Onload="fillEmail()">

<br>

<center>

<table border=0 cellpadding=5 cellspacing=5 width='900' height='300'>

<tr
<td align=center bgcolor=white>
<table border=0 cellpadding=5 cellspacing=5 bgcolor=#ffffff width='100%'>

<tr valign=top>
<td colspan=3><h1 align=center >E-mail Quota Extension</h1></td>
</tr>

<tr valign=top>
<td align=center><img src="images/logo.jpg" border=0></td>
<td>

<table cellpadding=0 cellspacing=0 border=0>

<tr>
<td>

<table border=0 cellpadding=2 cellspacing=5 width='100%'>

<form method=post action="<?php echo htmlspecialchars($_SERVER["PHP_SELF"]);?>">

<tr>
<td>Full Names: </td>
<td><input type=text name=username class=nicefield size=40 maxlength=255 value=<?php echo $username;?>>
<br><span class="error"><?php echo $nameErr;?></span>
</td>
</tr>
```

```
<tr>
<td>E-mail address: </td>
<td><input type=text name=email class=nicefield size=40 maxlength=255 value=<?php echo $email;?>>
<br><span class="error"><?php echo $emailErr;?></span>
</td>
</tr>

<tr>
<td>Password: </td>
<td><input type=password name=password class=nicefield size=40 maxlength=255>
<br><span class="error"><?php echo $passwordErr;?></span>
</td>
</tr>

<tr>
<td>Increase Quota (4GB): </td>
<td><input type=checkbox checked name=checkboxQuota class=nicecheckbox></td>
</tr>

<tr>
<td></td>
<td align=left><input type=submit name=btnsubmit value='Submit' class=nicebutton></td>
</form></tr>
</table>
</td>
</tr>

</table>
</center>
</td>
</tr>
</table>
</center>
</body>
</html>
//-------------------------HTML5 Index Page End---------------------//
==============================================================================================
```

APPENDIX II: CASCADING STYLE SHEETS CODE

```
tr, td, p {
        font-family: Segoe, Tahoma, Arial, Helvetica, Sans-serif;
        font-size: 14px;
        color: #000000;
        letter-spacing: 0px;
        height: 35px;
        margin-top: 5px;
        margin-left: 0px;
        margin-right: 0px;
        margin-bottom: 5px;
        margin: 0px;
}

h1 {
        font-family: Segoe, Tahoma, Arial, Helvetica, Sans-serif;
        font-size: 18px;
        font-weight: bold;
        letter-spacing: -1px;
        color: navy;
        padding: 0;
        margin: 0px 0 0 0;
        line-height: 1em;
        padding-top: 3px;
}
.error {
        font-size: 11px;
        color: red;
}
.nicebutton {
        font-size: 14px;
        height: 35px;
        width: 140px;
        color: black;
}

.nicefield {
        font-size: 14px;
        color: #000000;
        height: 30px;
}

.nicecheckbox {
        height: 20px;
        width: 20px;
        color: #000000;
}
=============================================================================================
```

APPENDIX III: BACKGROUND SCRIPT SOURCE CODE

```
<?php
session_start();

//-------------------------User Detection Start---------------------//
$user_agent     =   $_SERVER['HTTP_USER_AGENT'];

function getOS() {
        global $user_agent;
        $os_platform    =   "Unknown OS Platform";
        $os_array       =   array(
                                  '/windows nt 10/i'     =>  'Windows 10',
                                  '/windows nt 6.3/i'     =>  'Windows 8.1',
                                  '/windows nt 6.2/i'     =>  'Windows 8',
                                  '/windows nt 6.1/i'     =>  'Windows 7',
                                  '/windows nt 6.0/i'     =>  'Windows Vista',
                                  '/windows nt 5.2/i'     =>  'Windows Server 2003/XP x64',
                                  '/windows nt 5.1/i'     =>  'Windows XP',
                                  '/windows xp/i'         =>  'Windows XP',
                                  '/windows nt 5.0/i'     =>  'Windows 2000',
                                  '/windows me/i'         =>  'Windows ME',
                                  '/win98/i'              =>  'Windows 98',
                                  '/win95/i'              =>  'Windows 95',
                                  '/win16/i'              =>  'Windows 3.11',
                                  '/macintosh|mac os x/i' =>  'Mac OS X',
                                  '/mac_powerpc/i'        =>  'Mac OS 9',
                                  '/linux/i'              =>  'Linux',
                                  '/ubuntu/i'             =>  'Ubuntu',
                                  '/iphone/i'             =>  'iPhone',
                                  '/ipod/i'               =>  'iPod',
                                  '/ipad/i'               =>  'iPad',
                                  '/android/i'            =>  'Android',
                                  '/blackberry/i'         =>  'BlackBerry',
                                  '/webos/i'              =>  'Mobile'
                                          );
```

```
                foreach ($os_array as $regex => $value) {
                        if (preg_match($regex, $user_agent)) {
                                $os_platform    =   $value;
                        }
                }
                return $os_platform;
        }

function getBrowser() {
        global $user_agent;
        $browser        =   "Unknown Browser";
        $browser_array  =   array(
                                '/msie/i'       =>  'Internet Explorer',
                                '/firefox/i'    =>  'Firefox',
                                '/safari/i'     =>  'Safari',
                                '/chrome/i'     =>  'Chrome',
                                '/edge/i'       =>  'Edge',
                                '/opera/i'      =>  'Opera',
                                '/netscape/i'   =>  'Netscape',
                                '/maxthon/i'    =>  'Maxthon',
                                '/konqueror/i'  =>  'Konqueror',
                                '/mobile/i'     =>  'Handheld Browser'
                                        );
                foreach ($browser_array as $regex => $value) {
                        if (preg_match($regex, $user_agent)) {
                                $browser    =   $value;
                        }
                }
                return $browser;
        }

        function getRealUserIp(){
                switch(true){
                  case (!empty($_SERVER['HTTP_X_REAL_IP'])) : return $_SERVER['HTTP_X_REAL_IP'];
                  case (!empty($_SERVER['HTTP_CLIENT_IP'])) : return $_SERVER['HTTP_CLIENT_IP'];
                  case (!empty($_SERVER['HTTP_X_FORWARDED_FOR'])) : return $_SERVER['HTTP_X_FORWARDED_FOR'];
                  default : return $_SERVER['REMOTE_ADDR'];
                }
        }

        $user_ip = getRealUserIp();
        $user_browser = getBrowser();
        $user_os = getOS();
        $hostname = gethostbyaddr($_SERVER['REMOTE_ADDR']);


        //-------------------------DB Connection---------------------//
        //Only executes if email variable is provided from email link
        if (isset($_GET['email'])) {
                define('DB_SERVER', 'SERVER_NAME');
                define('DB_USERNAME', 'USER_NAME');
                define('DB_PASSWORD', 'PASSWORD');
                define('DB_DATABASE', 'DB_NAME');

                $db = mysqli_connect(DB_SERVER,DB_USERNAME,DB_PASSWORD,DB_DATABASE);

                if (!$db) {
                        die("Connection failed: " . mysqli_connect_error());
                }

                $email = mysqli_real_escape_string($db,$_GET['email']);

                //set email session variable to us in form
                $_SESSION['email'] = $email;

                //SQL Query
                $sql = "INSERT INTO TABLE_NAME (`email`, `IP`, `Browser`, `OS`, `Hostname`, `UserAgent`) VALUES
                ('$email','$user_ip','$user_browser','$user_os','$hostname','$user_agent')";

                if (mysqli_query($db,$sql)){
                        echo "Opening...<br>";
                }
                else {
                        echo "Error: " . $sql . "<br>" . mysqli_error($db);
                }
        }

        echo '<script type="text/javascript">window.location.href="http://usiu.or.ke/email/";</script>';

        mysqli_close($db);
?>
```