Understanding the Human Behavioural Factors behind Online Learners' Susceptibility to Phishing Attacks

Ayman Shargawi, MSc

November, 2017

This thesis is submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy.

Department of Educational Research,
Lancaster University, UK.

Understanding the Human Behavioural Factors behind Online Learners' Susceptibility to Phishing Attacks

Ayman Shargawi, MSc

Declaration

This thesis results entirely from my own work and has not been offered previously for any other degree or diploma.

The word count of this thesis is (51,565 words)

Signature ....

Abstract

Phishing is an act of fraudulence to lure victims to respond to an illegitimate request for the sake of a financial or informational gain (Huang, Qian, and Wang, 2012). Phishing can jeopardize the security of online learning (e-Learning) systems. Phishing cannot be prevented by depending on technical controls alone (Proctor, Schultz and Vu, 2009). Effective Information Security Awareness is key to protecting against Phishing (Chen, Shaw and Yang, 2006). However, most information security awareness programs overlook human behavioural factors as a root cause of exploitation in Phishing (Proctor et al 2009, Anttila et al 2007).

This research aims to better understand the human behavioural factors behind online learners' susceptibility to Phishing attacks (Luo et al, 2013). Thus, literature review was conducted to identify and analyse the human behavioural factors exploited in Phishing attacks with relation to the online learners' awareness needs. A conceptual framework called 'Security Awareness Model for Phishing' (SAMFP) has been developed based on the integration of Endsley's Situation Awareness model (Endsley, 2015), the awareness delivery guidelines by Chen, Shaw and Yang (2006) and Poepjes' (2012) Information Security Awareness and Capability Model (ISACM). SAMFP aims to improve information security awareness for online learners. Hence, data was gathered from 100 participants, experienced in learning online, who completed 5 activities: a pre-awareness (1^{st}) assessment test, participating in the 1^{st} awareness session and group discussions, an assessment (2^{nd}) test, participating in the 2^{nd} awareness session and group discussions and finally a post-awareness (3^{rd}) assessment test. Data was analysed quantitatively with 18 hypotheses to validate the effectiveness of the SAMFP model. Following a design based research approach, the researcher was heavily engaged in the design, development and testing of the SAMFP model which included development of training materials, tutoring and assessment of learning outcomes against the research questions and objectives.

Contents

List of Figures

List of Tables

Chapter 1 Introduction and Background

Thanks to the Internet and the evolution of web technologies with features such as collaboration, knowledge sharing, user-content generation, collective intelligence, creativity and innovation (Zuev, 2012), E-Learning systems such as Moodle and Blackboard have become an essential part of the modern society (Rjaibi, Rabai, Aissa, and Louadi, 2012). Public and private universities and institutions (Rjaibi et al, 2012; Alecu, Pocatilu and Capisizu, 2010) have adopted e-Learning systems as a credible means of education and training to share and distribute information electronically (Chen and He, 2013). Even in the business world, the need for online learning is on the rise (Chen and He, 2013).

The Internet has revolutionized the way people learn as it has brought great flexibility in terms of time and space into online learning. As such, public and private universities and institutions (Rjaibi et al, 2012; Alecu, Pocatilu and Capisizu, 2010) have adopted online learning as a credible means of education (Chen and He, 2013). Yet, it has introduced potential Information security risks due to its openness (Vasilescu, Tatar and Codreanu, 2011), diverse technologies and interconnectedness. One of the major information security risks introduced to Online Learning is Phishing. Phishing is a type of identity theft and an act of fraudulence to lure unsuspecting victims to believe that an illegitimate website or a link is legitimate for the sake of a financial or informational gain (Huang, Qian, and Wang, 2012; Zhang, Hong and Cranor, 2007). Phishing uses Social Engineering which involves fraudulent manipulation techniques to trick people into revealing sensitive information by posing as a trustworthy entity to them (Steyn, Kruger and Drevin, 2007; Parrish, Bailey and Courtney, 2009).

E-Learning Systems' stakeholders such as designers, developers, tutors, students, administrators and managers should be aware of such risks in order to play their role effectively in protecting the Confidentiality, Integrity and Availability (McCumber, 1991; Vasilescu, Tatar and Codreanu, 2011) of E-Learning systems. The diversity of technologies used coupled with the dependency on the Internet which is relatively insecure for communication (Barik Karforma, 2012; Costinela-Luninita and Nicoleta-Magdalena, 2012; Luminita, 2011; Rjaibi et al, 2012) have introduced a lot of vulnerabilities into E-Learning systems. These vulnerabilities need thoughtful adoption of appropriate security controls to mitigate their exploitation (Vasilescu, Tatar and Codreanu, 2011; Zuev, 2012). However, a major security control may be overlooked, that is users' security awareness.

A major Information Security Preventive control may have been overlooked, that is Online Students' and online tutors' awareness about Phishing and how Phishing attacks are schemed to target its victims. In the literature, most attention has been given to the risks of Phishing that could threaten the security of information, whereas little attention has been given to the issues of lack of awareness in research and practice (Vasilescu, Tatar and Codreanu, 2011; Aljawarneh, 2011; Yong, 2007) or to the reasons why victims fall for Phishing attacks. In other words, Online Students' and online tutors' lack of awareness about the behavioural factors, that are used as incentives to lure victims into Phishing traps such as: temptation (Jagatic, Johnson, Jakobsson, and Menczer, 2007; Kirlappos and Sasse, 2012), curiosity, urgency (Kirlappos and Sasse, 2012), threatening and over-trust (Kirlappos and Sasse, 2012; Furnell, 2008; Kumaraguru et al, 2010), deserves further attention.

In fact, few organizations including universities adopting Online Learning have security awareness programs in place. 67% of 400 information security officers described security awareness in their organizations as 'inadequate' as per 2002 Pentasafe Security Technologies report (Chen, Shaw and Yang, 2006). Despite all the sophisticated technologies used for protection, lack of security awareness can make an organization vulnerable to all kinds of external and internal threats (Chen, Shaw and Yang, 2006). Most of Information Security-related risks especially Phishing which is the focus of this research, are deeply interwoven with interaction with human behavioural factors. Therefore, Online Students' awareness about behavioural factors plays a major and overarching role in combination with technical countermeasures to effectively mitigate the risks of Phishing (Proctor, Schultz and Vu, 2009).

Despite the growing use of E-learning systems reported by many studies (Chen and He, 2013), little attention has been given to the issues of lack of information security in e-Learning systems' users (Vasilescu, Tatar and Codreanu, 2011; Aljawarneh, 2011; Yong, 2007). For this reason, this project aims to bridge this gap, research this area and find out how to improve Online learners' awareness about these behavioural factors underlying Phishing attacks. Therefore, a conceptual framework is needed to help improve Online learners' awareness about the behavioural factors underlying Phishing attacks. Based on the literature review conducted in chapter 2, a Situation Awareness model by Endsley (2015) has been identified to be a suitable conceptual framework to propose for this purpose thanks to its sustainability and dynamicity to meet the requirements for building measurable, robust and in-depth knowledge in participants about Phishing attacks exploiting behavioural factors over changeable space and time, integrated with a set of pedagogical guidelines recommended by Chen, Shaw and Yang (2006) and measured by the use of Poepjes' (2012) Information Security Awareness and Capability Model (ISACM) – more details can be found under the Theoretical Background section. The reason for the suitability of Endsley (2015) Situation Awareness Model lies in its dynamic use of time and space as measurements of situation change, which makes it capable of providing continuity and sustainability of awareness in a constantly changing field like information security.

1.1 Research Problem Statement

Phishing awareness mainly focused on visual cues and Internet protocol syntactical indicators to be identified in order to recognize Phishing attacks (Kirlappos and Sasse, 2012), despite the fact that many studies showed that unawareness of how some human behaviours are exploited and manipulated by Phishers to deceive victims played a key role in victims falling prey to Phishing attacks (Proctor, Schultz and Vu, 2009; Vasilescu, Tatar and Codreanu, 2011; Chen and He, 2013).

As the Internet has become the primary media for conducting e-learning for which many online learning providers have adopted technical controls such as anti-virus detectors and firewalls to protect their systems and information, yet the human related behavioural factors based awareness are neglected (Luo et al, 2013). According to Chen, Shaw and Yang (2006), the biggest challenge facing Information Security Awareness is the changing nature of ICT and the agility of applications. As a result, E-Learning systems required security awareness providers to use delivery methods that are more interactive and collaborative. This means incorporating activities to encourage interactive participation from users to help promote awareness among users while exchanging and sharing experiences. Oppositely, traditional training methods have been shown ineffective in raising awareness about Phishing risks. (Anttila et al, 2007). This is in addition to the fact that human behavioural related exploitation in Phishing attacks is still under-explored (Zuev, 2012) and therefore needs further research to uncover it. Therefore, understanding human behaviours and providing relevant security awareness sessions to address all kinds of exploitation of human behaviour factors such as demotivation, low self-esteem, technology incompetence, over-competence are very essential to constitute effective information security awareness programs to provide for information protection (Zuev, 2012).

1.2 Research Context

The stage has been set for this research to focus on the researcher and participants on how to better understand the human behavioural factors behind Phishing attacks targeting online learning users and systems. According to Parrish, Bailey and Courtney (2009), these human behavioural factors can be exploited as luring incentives in phishing attacks. Thus, the context of this research will address these behavioural factors and the online learners that use the Internet and the online learning systems for education as potentially susceptible targets for such Phishing attacks. The context will also address proposing and testing a conceptual framework for achieving higher levels of awareness in online learners about these behavioural factors and how they may be exploited in Phishing attacks.

Lack of awareness about Phishing human behavioural factors (Proctor, Schultz and Vu, 2009) threatens the security of e-Learning systems and would increase the susceptibility of online learners for falling for Phishing attacks if not given the right attention it deserves. For this reason and the significant risks emerging from this lack of awareness (Styne, Kruger and Drevin, 2007) and the pressing need for a dynamic and practical framework to improve online learners' awareness, this research has been contextualized to attempt to bridge this awareness gap and establish a frontline layer of defence against online learning based Phishing attacks.

This research was conducted with a focus on linking Information Security as the researcher's practitioner field with Educational Theory in general and TEL in particular to explore and investigate e-Learning users' susceptibility to Phishing attacks (The information security part) and to link that with Learning theories (TEL part) to explore ways to enhance the level of awareness about Phishing attacks targeting online learners. The collective research in both realms have contributed the conceptual framework SAMFP which is based on a combination of learning theories to address the awareness residual risk in online learners in particular and general online users in general. As such, this research has been conducted on 100 online learners, however, the results of the study could indirectly be applicable to any type of online users in general since all online users including e-learning users rely on the Internet and online environments to conduct their activities. Hence, they use the same medium through which online Phishing attacks can take place.

1.3 Research Questions

This research will attempt to answer the following questions:

1. What are the potential human behavioural factors exploited by Phishers to design and launch Phishing attacks against online learners?

    1a: How are these human behavioural factors exploited to lure targeted online learners to respond to Phishing attacks?

2. What is the preliminary level of online learners' awareness about these human behavioural factors and the ways they are used by Phishers to set up Phishing attacks?

    2a: How can online learners' awareness be improved and sustained as a frontline protection measure against these Phishing attacks?

1.4 Research Goals

The research aims to achieve the following goals:

- To better understand the human behavioural factors that are exploited by Phishers to attack online learners exploiting their lack of awareness about Phishing and the vulnerabilities of behavioural factors.
- To propose and test a conceptual framework that can help raise and improve online learners’ awareness to pose as a frontline protection measure against Phishing attacks.
- To assess the effectiveness of the proposed conceptual framework by measuring the learning outcomes and analysing the awareness improvement levels resulting from applying it.

1.5 Conceptual Framework

A conceptual framework is proposed and developed based on Endsley's (2015) Situation Awareness, the pedagogical guidelines of Chen, Shaw and Yang (2006) and Poepjes' (2012) Information Security Awareness and Capability Model (ISACM) to undertake the tasks of this Design-Based research. The conceptual framework is called 'Situation Awareness for Phishing Model' (SAMFP). The SAMFP model is used to achieve the goals of the research and eventually provide answers to the research's two questions by allowing the researcher and participants to better understand the human behavioural factors exploited in Phishing attacks through effective awareness methods that help participants build knowledge through engaging in interactive group discussions. The use of group discussions whether online or Face-to-Face as part of designing and implementing the conceptual framework has really revolutionized the way awareness sessions are delivered. By engaging participants in these group discussions, way was given to better learning results to emerge through interactivity and knowledge sharing. The three components incorporated in the SAMFP model helped it achieve these goals by a) adopting Endsley's (2015) Situational Awareness Model as its base framework to implant knowledge in the participants, b) adopting Chen's et al (2006) pedagogical guidelines to overcome awareness delivery issues realized from the one-way communication based awareness methods normally used in Information Security awareness and b) incorporating Poepjes' (2012) ISACM Model along with other devised measurement tools to measure the learning outcomes and awareness improvements achieved by participants after implementing the proposed SAMFP model.

1.6 Research Methodology

The implementation of the conceptual SAMFP model requires the researcher to design a solution, implement it and then assess it. The researcher's involvement in these three tasks made this research a Design Based Research (DBR) (Andriessen, 2007). This involvement is demonstrated by the researcher being the designer of a solution to the research problem through developing an instructional design of a full awareness program with all its resources and materials including the assessment tests. As such, putting the solution to test by applying the SAMFP model to a sampled population and finally assessing its effectiveness by measuring the resulting learning outcomes. Quantitative data analysis (Cohen, Manion and Morrison, 2011) is used to statistically assess and measure the significance of the outcomes from the implementation.

This research used convenience or non-probability sampling despite the argument in literature about using convenience sampling in social qualitative research being subject to the risk of selection bias (Etikan, 2016; Coyne, 1997). However, it is considered an acceptable alternative sampling approach when used with compensating precautions (Magnani, Sabin, Saidel and Heckathorn, 2005). Therefore, considering the fact that this research is a quantitative study that focuses on participants being online learners only, regardless of any other traits, combined with the hardship associated with participant recruitment within the timeframe allotted, this research used convenience sampling to select the minimum required number of participants (which is 100) who should complete the 5 required steps of participation in order for their data to be included in the analysis. Thus, the sampling approach taken in this research is convenience sampling requiring participants only to be online learners regardless of any other traits such as specific location, discipline or culture. Therefore, the distribution of participants in terms of location and culture was not part of the selection strategy or the data analysis. As such, this is a limitation of the research but could also be a potential opportunity to extend this in future research studying the effect of a particular sampling distribution that is based on specific demographics such as culture, nationality, location, education major or career on online learners' susceptibility to Phishing attacks.

This research project will undertake four steps to achieve the research’s goals: 1) Conduct a Literature Review to identify potentially exploitable human behavioural factors that pose a threat to online learning learners; and the frameworks to help better understand them. 2) Conduct a Pre-Awareness assessment test to assess the level of awareness that online learners have about these behavioural factors. 3) Conduct two Awareness sessions to participating online learners to help them improve their level of awareness about Phishing and how Phishing attackers exploit vulnerable human behaviours to trap victims. Finally 4) conduct a Post-Awareness assessment tests to measure the improvements made by the participants after their involvement in the awareness sessions.

1.7 Researcher’s Role and Impact

The role of the researcher in such Design-Based research is very dynamic. Part of the researcher’s knowledge strengthening engagements in the research was to practice teaching online and in classroom conducting awareness. This was a fruitful experience that located the researcher in the heart of the action. This had allowed the researcher to observe participants, their reactions and their learning progress more closely and with in-depth understanding of how situations developed and redeveloped according to the changing parameters of time and space e.g. after different elapses of times and in learning environments such as online and classroom based. At the same time, caution was always taken by the researcher not to influence participants’ learning and judgement by not showing bias to any actions to drive participation into the researcher’s favoured direction. Observations were recorded as they happened to ensure the reliability of the data collected for analysis.

1.8 Contributions of this Research

The conducted research linking both realms Information Security and technology-enhanced learning (TEL) have contributed the conceptual framework SAMFP (see chapter 3) which is based on a combination of learning theories to address the awareness residual risk in online learners in particular and online users in general. As such, the outcome of the study could indirectly be applicable to any type of online user in general since all online users including e-learning users rely on the Internet and online environments which are venues for online Phishing attacks, to conduct their activities.

The research contributed to knowledge with a proof of concept about the effectiveness of the proposed conceptual framework namely Situation Awareness for Phishing (SAMFP) that is derived from Endsley's Situation Model (2015), the pedagogical guidelines of Chen, Shaw and Yang (2006) and Poepjes' (2012) Information Security Awareness and Capability Model (ISACM); to improve awareness in online learners about 16 human behavioural factors exploited in Phishing attacks and measure the outcomes of such awareness in a continuous and iterative manner. The contributed proof of concept provides the evidence in the form of 18 hypotheses statistically tested yielding significant findings that the conceptual SAMFP model and the process designed for implementing it were demonstrated to be effective for improving the awareness level of a 100 online learners about the identified 16 human behavioural factors related to Phishing. Furthermore, the feedback from participants about their reasons for considering the given scenarios on the 1st Test as Phishing had also contributed with novel knowledge about the participants' preliminary understanding about Phishing and the related 16 behavioural factors.

The SAMFP model was a novel contribution to the theory of providing effective cyber security awareness about Phishing, while the implementation of the model was a contribution to the practice of conducting cyber security awareness about Phishing to online learners. Additionally, the results obtained from that implementation could contribute to making an effective policy for providing and conducting cyber security awareness about Phishing within an organization.

Additionally, the literature review contributed with the identification of the 16 behavioural factors and the ways of exploiting them in Online Learning based Phishing attacks. All the studies referenced in the literature review had studied only some of the behavioural factors related to Phishing based on general scenarios as opposed to online learning scenarios in particular. Nonetheless, not a single research was found to have covered all the 16 identified behavioural factors and thoroughly studied their exploitation methods in Online Learning based Phishing attacks, nor thoroughly had attempted to explore and improve online learners' awareness and understanding about them.

Furthermore, the implementation of the SAMFP model made a novel contribution by identifying the preliminary awareness level of 100 online learners about Phishing and these 16 behavioural factors exploited in Phishing attacks. This contribution was considered as a set point for further iterations of awareness sessions and assessment tests to start by using the SAMFP model to enhance the 100 participants' preliminary awareness level to higher levels of awareness.

In summary, this research contributed by 1) identifying 16 Phishing related behavioural factors, 2) a snapshot of current 100 online learners' awareness about these Phishing behavioural factors, 3) a novel conceptual framework model called SAMFP to improve information security awareness and 4) providing evidence of the efficacy of the SAMFP model by statistically showing it has helped improve the awareness of 100 online learners.

The contributions of this research are intended for information security practitioners who are tasked with providing effective cyber security awareness about Phishing to their audiences and organizations, Information security and educational researchers who are interested in experimenting with different methods of conducting awareness and finding ways to improve the learning outcomes from cyber security awareness and finally the cyber security awareness free-lancers who conduct online and classroom based cyber security awareness sessions to the public with the aim of enhancing their level of awareness about cyber security.

1.9 Thesis Overview

In this introductory chapter, the research problem statement, research questions and goals are discussed. Also, a brief summary of the research conceptual framework, methodology, role of the researcher and research contributions are provided. The thesis consists of the following chapters:

1. Literature Review: Discusses findings resulting from literature review namely, identifying the 16 behavioural factors potentially exploited in Phishing attacks and the conceptual frameworks to help improve awareness about it.

2. Research Design: The research methodology and methods are discussed here. Also, a detailed discussion of the conceptual framework and the tasks required to conduct the research is provided.
3. Data Collection: The data collection strategy and approach, methods, tools and issues are all discussed.
4. Data Analysis: The data analysis methods, tools and tests are discussed and explained in detail supported by examples from the data collected. Also, the hypotheses set and their analysis are thoroughly discussed and followed by a discussion of correlations between the results of the analysis findings and literature review. In addition, the recruitment of participants and sampling are discussed.
5. Results: This chapter explains the results of the data analysis findings.
6. Results Discussion, Contributions, Implications and Future Research: This chapter discusses how the results relate to the research questions and objectives; followed by a discussion of the implications of this research and the avenues it opens for future research and finally discusses the contributions of the research.

Chapter 2 Literature Review

This literature review has been conducted to explore the existing body of knowledge surrounding the subject of this research which is better understanding Phishing attacks and enhancing awareness about them in online learners to provide a solid background supporting the goals set for the research. In parallel, this review is also part of gathering the required information about any potential human behavioural factors that are exploitable in phishing attacks as this information will provide the answer to the first part of the 1st question of this research. The following areas were covered as part of this review: Online learning and phishing, e-learning systems and online learning, information security in online learning systems, awareness in secure online learning, Information Security Awareness Tools, role of ISS support to protect against phishing, and behavioural factors employed to deceive victims in phishing attacks.

2.1 Online Learning and Phishing

Online Learning depends on the Internet and the use of e-mail to deliver content; which make it vulnerable to Information Security risks such as Phishing as any other business organizations (Steyn, Kruger and Drevin, 2007). One of the major information security risks introduced in Online Learning is Phishing. Phishing is a type of identity theft and an act of fraudulence to lure unsuspecting victims to believe that an illegitimate website or link is legitimate for the sake of a financial or informational gain (Huang, Qian, and Wang, 2012; Zhang, Hong and Cranor, 2007). Another definition of Phishing is that it is a type of a cyber-attack using Social Engineering techniques which involves fraud to trick people into revealing sensitive information by posing as a trustworthy entity to them (Steyn, Kruger and Drevin, 2007; Parrish, Bailey and Courtney, 2009). An example of a Phishing case was when hackers used social engineering techniques to trick Ford credit employees into revealing an authorization code that was used to access the personal information of about 13,000 individuals who were exposed to the risk of identity theft (Proctor, Schultz and Vu, 2009).

Information Security technical controls could be rendered useless without proper understanding of the risks resulting from the lack of awareness and irresponsible behaviour. As such, the lack of awareness in online learners about the security risks associated with Online Learning systems is a source of a vulnerability that could be exploited by insiders and outsiders alike. Attackers would conduct a Phishing attack on a target by manipulating him/her to break into the system and cause damage (Chen and He, 2013) and therefore, jeopardize the confidentiality, integrity and availability of the system and its resources. Most internal abuse is caused by users' lack of awareness where the human behaviour factor is more vulnerable to exploitation than technology (Vasilescu, Tatar and Codreanu, 2011).

Unfortunately, Chen and He (2013) stated that many organizations are rushing the use of online learning without carefully thinking about the security threats and risks inherent to it. Accordingly, a study was carried out to test the level of awareness in an academic environment about Identity theft revealed that more than 50% of respondents gave away their passwords in response to a phishing test email (Steyn, Kruger and Drevin, 2007). According to another study in 2004 by the Gartner group, about 57 million US adults received a phishing email whereas almost 11 million online adults have clicked on a link in phishing emails (Steyn, Kruger and Drevin, 2007). According to a third study conducted jointly by the Computer Security Institute and the Federal Bureau of Investigation (FBI) in 2005, Phishing attacks are the most common type of electronic thefts and fraud. More than 260,000 cases of identity fraud were identified in the study (Chen, Shaw and Yang, 2006). Next, Online Learning Systems are discussed.

2.2 E-Learning Systems and Online Learning

The first general-purpose E-Learning system ever developed was the Programmed Logic for Automatic Teaching Operations (PLATO) by the Computer-based Education Research Laboratory (CERL) in 1960 (Basha and Dhavachelvan, 2010). Accordingly, E-Learning in general refers to computer-enhanced learning or in other words, it can be described as doing learning activities electronically through the Internet (Barik and Karforma, 2012). Also, Chen and He (2013) described E-Learning as using content repositories to store content while facilitating communication through web-based technologies that learners can interact with to collaboratively learn and share knowledge.

Therefore, Online Learning depends on the Internet to execute and so do E-Learning Systems. The use of e-mail and the Internet at academia makes it as vulnerable to Phishing as any other business organizations (Steyn, Kruger and Drevin, 2007). Technical controls are of less value without proper understanding by the workforce of the risks of their irresponsible behaviour. Therefore, they inherit all the security risks associated with the use of the Internet. Hence, information security awareness becomes necessary to enhance the posture of information security in online learning systems and environments through raising awareness of online learners about the risks associated with their roles in the information security process (Kruger et al, 2008)

Information Security depends on three categories in the analysis of security threats and risks. These three categories make the Information Security triad (CIA) which refers to Confidentiality, Integrity, and Availability (McCumber, 1991; Vasilescu, Tatar and Codreanu, 2011). Accordingly, inherent threats of the Internet (Aljawarneh, 2011) and the diversity of technologies underlying E-Learning systems may lead to the exploitation of vulnerabilities in these three areas (CIA) creating potential risks (Chen and He, 2013; Zuev, 2012; Lim and Jin, 2006). Below are some more examples of risks under each area of the Information Security Triad (Costinela-Luninita and Nicoleta-Magdalena, 2012; Luminita, 2011; Rjaibi et al, 2012):

- Confidentiality Risks: Such as data leakage.
- Integrity Risks: Such as cross site request forgery and scripting.
- Availability Risks: Such as Denial of Service attacks.

A security threat can be any anticipated danger that may exploit a vulnerability which is a weakness or a loophole in a system producing risk which is the probability of a particular threat exploiting a vulnerability (Barik Karforma, 2012). Thus, a security breach can be defined as an event where a vulnerability in a system is exploited to subvert a security control to commit harmful acts (Proctor, Schultz and Vu, 2009). Information Security's ultimate goal is to protect the confidentiality, integrity and availability (CIA) of information (McCumber, 1991; Vasilescu, Tatar and Codreanu, 2011). One of the major threats that can impact upon the confidentiality, integrity and availability of online eLearning systems is Phishing. As such, Phishing can jeopardize the confidentiality, availability and integrity of online eLearning systems and the information processed within them.

To combat the risks of threats such as Phishing, many online learning providers have adopted technical controls such as anti-virus detectors and firewalls to protect their systems and information while controls based on the human related behavioural factors such as awareness were overlooked and neglected (Luo et al, 2013). The lack of users' awareness about the security risks and the protection methods is still a source of a weakness. Most e-learning systems were designed with more emphasis on functionality than on security. That was because E-Learning systems were built to introduce the culture of openness to learning as opposed to the restrictions of its brick-and-mortar counterpart in terms of space and time. As such, incorporating security in the design would highly limit the achievement of openness in E-Learning systems (Vasilescu, Tatar and Codreanu, 2011).

Information Security threats can originate from outside and inside of an organization whereas countermeasures can be preventive, detective, deterrent and recovery depending on the level of protection they are placed at (Chen, Shaw and Yang, 2006). As such, detective controls take over when preventive controls are bypassed while recovery controls take over when all other controls have failed. Therefore, E-Learning systems should be designed to fend off not only external threats, but also those threats coming from within such as the lack of users' security awareness (Costinela-Luninita and Nicoleta-Magdalena, 2012; Luminita, 2011). There is no absolute 100% security for any system, but carefully implemented security controls could minimize the possibility of risk occurrence (Lim and Jin, 2006). However, awareness can play a major role in minimizing risks for online learning systems. The National Institute of Standards and Technology (NIST) define information security awareness as a way of simply focusing individuals' attention on security concerns and how to respond accordingly (Chen, Shaw and Yang, 2006). This definition emphasizes the role of awareness as a method to instil security-aware culture within organizations changing online learning systems' users' behaviours and enforcing good security practices (Aljawarneh, 2011; Chen, Shaw and Yang, 2006). According to the study conducted by Chen, Shaw and Yang (2006) to investigate the awareness needs of an insurance company that has an e-business presence, the study revealed that implementing a good security awareness program requires incorporating activities to encourage interactive participation from users to help promote awareness among users as they exchange and share experiences, history of security breaches and lessons learned (Chen, Shaw and Yang, 2006). Most internal abuse is caused by users' lack of awareness where the human behaviour factor is more important than technology (Vasilescu, Tatar and Codreanu, 2011). Therefore, understanding human behaviours and providing relevant security awareness sessions to address all kinds of human vulnerabilities such as, but not limited to, demotivation, low self-esteem, technology incompetence, over-competence, lack of interest, lack of awareness, lack of responsibility and accountability and poor management are very essential to information protection (Zuev, 2012). Understanding the factors especially those that are human related which pose as luring incentives behind falling prey to Phishing attacks is under-studied and requires in-depth research to better understand them. In addition, an online leaner in particular as a subject of Phishing attacks is also under-explored. For these two reasons, this research has attempted to better understand those human related luring motives behind online learners’ susceptibility for falling victim to Phishing attacks and how their level of awareness about Phishing and its underlying factors can be raised and sustained as a frontline countermeasure against such attacks. Next, the Information Security in Online Learning Systems and its challenges are discussed.

2.3 Information Security in Online Learning Systems

The modern learning theory for information security awareness emphasizes the need for collective, interactive and collaborative networked communities to facilitate knowledge and experience sharing among all members whereupon new knowledge can be constructed collectively (Anttila et al, 2007). Knowledge construction goes through Six (6) layers of mental processes namely: 1) Knowing, 2) Comprehension, 3) Application, 4) Analysis, 5) Synthesis and 6) Evaluation. Security Awareness should take into consideration each phase of the knowledge construction process when designing awareness and selecting tools for that. Web 2.0 tools such as wikis, discussion forums and blogs have proven effective in accommodating such learning needs (Anttila et al, 2007). Vasilescu, Tatar and Codreanu (2011) see the goal of implementing information security for E-Learning systems is to achieve the following information security objectives:

- Confidentiality of content and users' data
- Integrity of content, tools and teachers and students data processed by the system against malicious acts.
- Availability of e-learning services
- Identification and Authentication to eliminate identity thefts and impersonation
- Authorization to limit access as needed to information to prevent data loss and leakage and unauthorized modification of information.
- Accountability whereby every transaction can be traced back to its initiator.

This goes in line with the goal of implementing information security in online learning in general which is meant for the protection from malicious or accidental misuse of resources (Chen and He, 2013). To do that, a robust security policy must be established to cover all potential risks and vulnerabilities in E-Learning systems including hardware, software, human resource and nature (Zuev, 2012). Second, organizations should conduct risk assessment to identify risks and prioritize them based on impact and develop an effective mitigation plan to mitigate each risk (Rjaibi et al, 2012). As seen in the literature review that most technical controls could be rendered useless due to the lack of effective security awareness. Information Security Awareness is as much culture as a corporate policy directive. Successful information protection strategies rely on information security awareness that focuses on culture and tries to instil the security sense in users' practices and behaviours (Aljawarneh, 2011; Chen, Shaw and Yang, 2006). Security in E-learning is based on policy and technology where policy states how and what technology is used while technology assists policy by providing best practices in information protection. However, policy and technology cannot work alone without the support of people (stakeholders' awareness of security risks and mitigation methods) to achieve security goals such as securing e-exams, eliminating cheating and resolving the identity verification issues in online assessment (Chen and He, 2013; Vasilescu, Tatar and Codreanu, 2011). Therefore, all society members need a level of awareness about information security risks according to their interaction level with the system (Anttila et al, 2007). While, end users need a basic level of awareness that qualify them to use computer networks, deal with critical information and authentication, be aware of cyberspace rules and have self-protection in privacy and sensitive situations, system developers of Information and Communication Technology (ICT) systems need awareness in more depth to be able to identify risks and apply appropriate security measures to mitigate them (Anttila et al, 2007). On the other hand, Security Managers should have the competence to manage their organizations to implement security in light of the security policy in place and motivate employees to comply accordingly (Anttila et al, 2007).

Although studies showed that many security and privacy measures are dependent upon human behaviours for being effective, not much research has been conducted to investigate the role of human behaviour factors in enabling security and privacy measures (Proctor, Schultz and Vu, 2009). As such, all stakeholders of E-Learning Systems from Administrators to Developers should be aware of this relationship and consider human behaviours in their designs (Anttila et al, 2007). However, users including online learners will not cooperate with these security systems unless they find them user-friendly and intuitive (Proctor, Schultz and Vu, 2009). Anttila, Savola, Kajava, Lindfors and Röning (2007) suggest that Information Security professionals at an organization should stay current on any changes to information security industry standards in order to be able to provide awareness to others on how to apply and conform to approved industry Information Security standards and guidelines. The International Organization for Standardization (ISO) 17799 security standard recommends that user security awareness should focus on Ten (10) major security domains (Chen, Shaw and Yang, 2006):

- Information Security Policy
- System Access Control
- System Development and Maintenance
- Personnel Security
- Physical and Environmental Security
- Security Organization
- Asset Classification and Control
- Communications and Operations
- Business continuity Management
- Compliance.

In the same way, Barroca and Gimenes, (2012) suggest that users who use social media and Web 2.0 tools should be aware of the potential security issues of privacy that are publicly available through dedicated awareness sessions. Security awareness sessions should address all kinds of human behaviours' vulnerabilities such as, but not limited to, demotivation, low self-esteem, technology incompetence, over-competence, lack of interest, lack of awareness, lack of responsibility and accountability and poor management (Zuev, 2012). One issue worth mentioning in the discussion of human behaviours-based awareness is the Online Trust (Beldad, Jong and Steehouder, 2010). Beldad, Jong and Steehouder (2012) see that people who are highly proficient with the web are more likely to have lower perceptions of the risks using the web and be inclined to trust online transactions and therefore need awareness to raise their sense of online potential security risks.

The biggest challenge facing Information Security Awareness is the changing nature of ICT and the agility of applications such as E-Learning systems which created a necessity for security awareness providers to use delivery methods that are more interactive and collaborative given the need for deepened security awareness for which traditional training methods have been proven ineffective (Anttila et al, 2007). Following are some challenges that Information Security Awareness programs need to address and focus on in order to instil security awareness into E-Learning Systems' stakeholders' culture.

Users should be aware of risks such as generating easy-to-guess passwords, while System Administrators should be aware of how to enforce complexity schemes to passwords. For example, a recommended awareness session should be given to educate users on how to use a passphrase as opposed to a password whereby the user generates a phrase and composes the password from the first initial letter of each word in the phrase after applying some character substitutions with digits/symbols where applicable such as converting the 'a' to an '@' and so on. Analysis of one password cracker revealed that 62% of passwords which did not contain a symbol or a digit were cracked as opposed to only 2% of passwords containing a symbol or a digit (Proctor, Schultz and Vu, 2009).

Users should be aware of how to determine when to reveal personal information in order to complete a transaction. Studies show that this practice could lead to identity theft (Proctor, Schultz and Vu, 2009). Users should be aware of how to read and understand web site privacy agreements and software certificates in order to securely accept them. Accordingly, system developers should be educated on how to make these technical documents more user-friendly and automated (Proctor, Schultz and Vu, 2009). Users should be aware of how to identify content that they should trust or ignore as this could be a type of Phishing and Social Engineering attacks (Proctor, Schultz and Vu, 2009). The common factor among all the above challenges is the requirement for raising awareness about information security risks for online learning systems' users. Next, the role of Information Security Awareness in Online Learning is discussed.

2.4 Awareness in Online Learning

Parrish, Bailey and Courtney (2009) studied the susceptibility of people to Phishing attacks and the impact of personal traits such as age and gender in correlation with the Big five (5) personality traits which are Neuroticism, Extraversion, and Openness to experience, Agreeableness, and Conscientiousness to explain why some people are more susceptible to Phishing attacks than others. The study revealed that understanding the impact of the 5 personality traits has a significant impact on predicting human behaviours and their relative susceptibility to Phishing. Likewise, a similar study was conducted by Schrammel, Köffel and Tscheligi (2009) to investigate the relationships between personality traits and information disclosure in online communities. A similar study by Wright and Marret (2010) aimed to better understand the behavioural factors that affect one's susceptibility to Phishing attacks. The study revealed that four behavioural factors are responsible for people's susceptibility to Phishing attacks and suggested that awareness about these behavioural factors should be incorporated in anti-phishing efforts.

Phishing awareness mainly focused on visual cues and indicators to identify and recognize Phishing attacks (Kirlappos and Sasse, 2012) using techniques such as online games (Kumaraguru et al, 2010) whereas, Phishers tend to exploit human vulnerabilities by tapping into behavioural factors such as temptation (Jagatic et al, 2007; Kirlappos and Sasse, 2012), curiosity, urgency (Kirlappos and Sasse, 2012), threatening and over-trust (Kirlappos and Sasse, 2012; Furnell, 2008; Kumaraguru et al, 2010) and using them as incentives to lure victims into their Phishing attacks. These behaviour factors are not well-covered by awareness programs. Therefore, understanding human behaviours and providing relevant security awareness sessions to address all kinds of human behavioural vulnerabilities is very essential to information protection (Zuev, 2012).

The role of security awareness is very important in supporting the overall security posture in Online Learning environments and should address all needs of stakeholders (Anttila et al, 2007) such as Online learners. In addition, Anttila et al (2007) also suggest the use of Formal Learning, Experience Gaming, Mentorship, Performance Support, Self-Learning, Community Based Learning and Information Learning as effective methods to support collective learning and the creation of information security-aware culture. Chen, Shaw and Yang (2006) suggest that the most effective way to educate Online Learning stakeholders is the situational user-centred learning approach which best suits online information security. Furthermore, Chen, Shaw and Yang (2006) recommend a set of guidelines to enable Information Security professionals to design effective multi-faceted security awareness programs that can address Online Students:

- Use two-way communication as opposed to one-way whereby users can interactively participate.
- Create measureable targets to assess awareness program outcomes before and after.
- Create flexible Awareness programs that can be modified as needed or as informed by their outcomes assessment.
- Make reachable programs by introducing diverse methods of communication such as e-mail, wikis, blogs, discussion forums, online surveys and Face-to-Face presentations.

- Support decision making by creating a repository to document past security incidents, events and lessons learned from previous awareness activities.

Online Learning and the Internet have become an essential part of our modern educational and business systems. Likewise, Information Security threats have. In order to combat those security threats such as Phishing, deploying technical state-of-the-art controls like setting up a firewall and powerful anti-virus software are not enough without sustaining a proper level of Information Security Awareness in people who use these systems. As suggested by Chen, Shaw and Yang (2006) the Situational Awareness user-cantered model is the most effective way to educate online students. Hence, the Situation Awareness Model (Endsley, 2015) is used to underlie this research's proposed conceptual framework and integrate the set of guidelines recommended by Chen, Shaw and Yang (2006) in its fabric to raise awareness in online learners about Phishing behavioural factors. In addition to Endsley's Situation Awareness framework (2015) and the guidelines of Chen, Shaw and Yang (2006), Poepjes' (2012) Information Security Awareness and Capability Model (ISACM) is also integrated as part of the proposed conceptual model for the sake of Phishing awareness raising and measurement. This model is called 'Security Awareness Model for Phishing' (SAMFP). Poepjes' (2012) Information Security Awareness and Capability Model (ISACM) is also based on Ensdley's Situation Awareness theory (2012) by capitalizing on the integration of three attributes of situational awareness to measure the awareness level achieved. These attributes are (1) the Awareness Importance, (2) the Awareness Capability which is the participants' knowledge or existing level of awareness about the risk and the required controls to prevent it and finally (3) the Awareness Risk which is the difference between the Awareness Importance and Awareness Capability. More on these integrations between the three components that formulate the proposed conceptual framework SAMFP is going to be covered in more detail in Chapter 3.

Most risks and vulnerabilities associated with E-learning Systems are protected against using technical controls. On the other hand, awareness can play a major role in protecting E-Learning systems just as much. According to Barik and Carforma (2012), risks in E-learning systems are associated with the 5 main E-Learning systems' stakeholders:

- Authors' (Course Developer) Risks such as unauthorized change of content
- Teachers' Risks such as cheating on assessment and availability of content.
- Managers' (System Admin) Risks such as identity theft, copyright infringement etc.
- System Developers' Risks such as poor encryption or clear-text password transmission or insecure code practices in general.
- Students' Risks such as misuse of credentials, phishing and lack of security awareness in general.

There is no absolute 100% security for any system, but carefully implemented security controls can minimize the possibility of risk occurrence (Lim and Jin, 2006). Therefore, a balance should be stricken to achieve the goals of security and usability. Understanding how online learners develop learning and awareness of the human behavioural factors is under-studied and requires in-depth research. In addition, an online learners in particular as a subject of Phishing attacks is also under-explored whereas academics in general have been the subject of Phishing related research. For these two reasons, this research is going to attempt to improve understanding of how online learners can enhance their level of awareness about Phishing and the human behavioural factors underlying Phishing attacks. Next, a review of existing information security awareness tools are discussed and contrasted to the proposed SAMFP model.

2.5 Discussion of Existing Information Security Awareness Tools

Most information security training programs fail because they lack interactivity and adventure with the audience. According to Kumaraguru (2009), this lack of interactivity in traditional cyber-security awareness is attributed to three reasons: 1) thinking that users will act proactively and seek training, 2) expecting users to have some knowledge about the content e.g. Phishing attacks and 3) training material has not been designed with learning principles in mind. Therefore, Information Security programs need to engage users by making them think and apply cyber-security concepts. This approach to cyber-security training is called 'Embedded Training' by which training material is integrated with day-to-day tasks. PhishGuru is an example of using this approach where users interact with training material while performing their tasks (Kumaraguru, 2009). Another example of this approach is using game-based tools such as 'Anti-Phish Phil' (Kumaraguru, 2009) for delivering information security awareness in an interactive and adventurous way. Preliminary results of using these game-based tools are encouraging as opposed to traditional cyber-security awareness that is distributed through e-mail or published online (Cone, Irvine, Thompson and Nguyen, 2007). Another anti-Phishing game-based tool that was developed is 'Phree of Phish' which provides cues to identify Phishing e-mails and URLs (Pars, 2017). This game was used to measure users' awareness improvements before and after training using a control group and an experimental group. The results showed a significant difference in the level of awareness gained by the experimental group after being trained using the game (Pars, 2017). However, the content focused on visual cues more than human behaviours in those anti-phishing game-based tools. Another game-based tool that is used to educate users about Phishing uses simulated sequential messages that includes a disguised Phishing e-mail in the thread of e-mails it sends to random users (Higbee, Belani and Greaux, 2013). The approach of phasing out the attack over sequential messages enable the users to understand the phishing techniques used. However, most of these messages that are crafted by the tool rely on the trusted contact factor to deceive the unsuspecting users. A similar tool based on simulated messages is PhishMe (PhishMe.com). PhishMe provides Learning by Doing to its users in order to educate them how to recognize Phishing attacks. It uses customizable human behaviour driven scenarios that resemble the organization's specific threat vector (PhishMe.com). GoPhish is also a platform that provides Phishing awareness through repeatedly practicing simulated Phishing attacks (GoPhish.com). On the other hand, Phishing IQ tests are used to educate users about Phishing through asking them questions. According to a study conducted by Anandpara, Dingman, Jakobsson, Liu and Rointestad (2007), these IQ tests are claimed to have failed to measure users' awareness and ability to recognize Phishing attacks. Phishing IQ tests had rather measure users' fear of Phishing (Anandpara et al, 2007). Finally, mindfulness techniques were also used as an approach to train users how to dynamically allocate attention to recognize suspicious content based on rule-based training (Jensen, Dinger, Wright, and Thatcher, 2017).

Next, the role of Information Security Services (ISS) in raising awareness about Phishing is discussed.

2.6 Role of ISS Support to Protect Against Phishing

The literature review cannot be concluded without talking about the role of Information Security Services (ISS) in protecting online learning systems from information security risks including Phishing. The role of organizations' ISS Support in protecting against Phishing focuses on implementing technical controls such as setting up firewalls, Intrusion Detection/Prevention systems and monitoring logs (Vasilescu, Tatar and Codreanu, 2011; Kruger et al, 2008). However, providing awareness about Phishing to users comes next on their list, and in some cases is inadequate or superficial mostly depending on learning visual clues such as grammar and spelling mistakes to identify phishing e-mails. This simple awareness given by ISS does not help users to understand the real behavioural factors behind victims falling prey to Phishing attacks which is what this study has explored and tested to strengthen participants' understanding and vigilance about Phishing behavioural factors. In addition, findings from the literature review about the limitations of the ISS role in awareness, are further explored by asking participants for feedback on the role of their own ISS in providing awareness. That has been achieved by asking the participants to evaluate their ISS on their role in supporting awareness on a scale of 5. The analysis of participants' evaluation of ISS informed the researcher how effective and supportive their role was in raising awareness among online learners. Next, the behavioural factors identified in the literature to have been exploitable in Phishing attacks are discussed.

2.7 Behavioural Factors Exploited to Deceive Victims in Phishing Attacks

Kirlappos and Sasse (2012) studied 36 online shoppers to explore their tendency to fall prey to Phishing on online shopping websites and how their misconceptions about secure Internet browsing would affect their behaviour online. It was revealed that most participants were led by their real-world experiences based on their level of awareness being limited to identifying visual security indicators in a website such as using a secure protocol denoted by a lock, a trusted logo or link or the overwhelming information provided on the website for them; to trust a website. They were completely oblivious of the real drivers by which they were internally vulnerable to such attacks such as the temptation of a first-come-first-serve voucher. The study found that temptation and the false sense of urgency were exploited to deceive online shoppers to trust a website. Kirlappos and Sasse (2012) suggested that information security awareness has to make a paradigm shift by focusing more on the dangers of vulnerable behavioural factors of online users than visual website cues and indicators.

Online users' disposition to trust, urgency and similarity are also other behavioural factors that were studied by Wright and Marett (2010) as hypothesized incentives to increases online users' susceptibility to Phishing deception as opposed to, disposition to suspicion and risk perception which are adverse behavioural factors to decrease online users' susceptibility to phishing deception. However, the findings from their study did not support the hypotheses made about the adverse behavioural factors decreasing the likelihood of deception.

Another study by Vishwanath, Herath, Chen, Wang, and Rao (2011) tested the susceptibility to Phishing of a university's real phishing attack's victims and their abilities for cognitive elaboration of Phishing cues to detect phishing e-mails. According to the study, urgency and e-mail volume overloading were supported by the study results to be the behavioural factors among others including physical cues that are highly contributing to victims' susceptibility to Phishing.

Luo et al (2011) introduced a theoretical framework based on a heuristic systematic model to investigate the human behavioural factors underlying victimization by Phishing. The theoretical framework was validated by an experiment that unveiled these underlying behavioural factors which are 1) urgency by which victims are pressured to take immediate action without allowing adequate cognitive systematic processing of the contextual environmental information surrounding the phishing message and 2) over-trust caused by a fake source of credibility using techniques such as 3) likability and similarity of genre using well-known message titles and logos, organizational mimicked communication styles by which victims are misled to trust.

Ibrahim (2016) studied the effect of personality traits on SMS Phishing Vulnerability identifying a number of human behavioural vulnerabilities behind Phishing attacks. Based on an investigation of Social Engineering Techniques conducted by Rusch (1999 as cited in Ibrahim, 2016), Ibrahim (2016) categorized these human behavioural vulnerabilities by three areas: emotions, attitude and belief. Emotions by which a phisher deceives victims through temptation, greed, excitement, urgency, threatening and fear. For example a phisher might tempt a victim by a false prize, using the scarcity of the prize and the limited offer time to pressurise the victim. An example of attitude and belief include behavioural factors such as over-trust, authority and reciprocation when a victim tends to trust a message just because it seems to be coming from a credible source or a source of authority such as a manager; which might be impersonated and falsified.

Similar to Luo et al’s study (2011), Williams, Beardmore and Joinson (2017) conducted a similar study using models such as the Elaboration Likelihood Model and the Heuristic-Systematic Model to investigate the effectiveness and influence of persuasion techniques used in Phishing attacks which exploit individual differences in human behaviours. These human behaviours are reciprocity e.g. when doing someone a favour in order to make him/her obliged to respond, conformity or social proof by which a victim is persuaded to conform to a certain action influenced by social norms and patterns, authority by which a victim is influenced to comply with certain action, urgency and scarcity by which a victim is pressured to take immediate action without considering consequences, and other human behaviour factors including dispositional trust, likability and similarity (Lederman, 2012), commitment and consistency, temptation, threatening and fear. Both models suggested that the effectiveness and influence of online persuasion techniques exploiting these human behavioural vulnerabilities depend on the depth and thoroughness of the recipient’s cognitive elaboration when processing the message.

Laribee et al (2006) designed trust and attack models that exhibit different ploys and tricks that exploit certain human behavioural factors to gain trust of the targeted victims in a coordinated iterative manner. The objective of their model is to formalize the social engineering attacks vector and infer good countermeasures to protect targets from such planned attacks. Among many human behavioural tactics included in the model is the diffusion of responsibility by which a victim is persuaded to take action deceived by a false feeling of lack of personal responsibility and a minimized feeling of guilt in that he or she will not be the only one held responsible for such an action.

Workman (2008) conducted an empirical study to synthesize theory from the marketing literature explaining consumer behaviour and the factors accounted for successful marketing techniques and investigate whether the same human behavioural factors may account for successful phishing attacks. Among the human behavioural factors studied that are commonly used for persuasion in marketing and are also suitable for use in social engineering are reciprocation, consistency and commitment, social proof, likability and credibility, threatening and fear, urgency and scarcity and interpersonal relationships. While some phishers tend to use temptation to dupe online users into divulging information, others may resort to developing interpersonal relationships with the victim to gain their trust and commitment to reveal confidential information or commit an action.

According to Lazy User Theory in Adams' (2012) study of mutual authentication and Phishing, a user would only do the minimum to satisfy his/her requirement taking the more convenient path to achieve that requirement. This user behavioural tendency for convenience and laziness could be exploited by Phishing attackers to tempt the user to take an insecure shortcut that would involve bypassing security checks rather than the secure path that would take more effort to fulfil their informational needs.

Finally, Hanamura, Takemura and Komatsu (2013) concluded their analysis of the characteristics of cyber-security victims represented by Japan Home Internet users that users' self-consciousness and over-confidence in their information security knowledge increased the probability of falling victim to Phishing attacks.

Based on this literature review of human behavioural factors exploited as a bait in Phishing attacks, the following 16 human behavioural factors are identified to underlie Phishing attacks and therefore are considered to construct the content of the three tests and the materials for conducting the 2 awareness sessions for this research:

1. Temptation e.g. greed etc. (Jagatic et al, 2007; Kirlappos and Sasse, 2012; Ibrahim 2016; Williams, Beardmore and Joinson, 2017),
2. Urgency or Scarcity (Kirlappos and Sasse, 2012; Vishwanath et al, 2011; Luo et all, 2011; Williams, Beardmore and Joinson, 2017; Workman, 2008),

3. Over-confidence or Self-Consciousness (Hanamura, Takemura and Komatsu (2013),
4. Dispositional (Over) Trust (Kirlappos and Sasse, 2012; Furnell, 2008; Kumaraguru et al, 2010; Wright and Marett 2010; Williams, Beardmore and Joinson, 2017; Hong, Kelley, Tembe, Murphy-Hill, and Mayhorn, 2013),
5. Authority (Ibrahim, 2016),
6. Threatening, Fear or Anxiety (Kirlappos and Sasse, 2012; Furnell, 2008; Kumaraguru et al, 2010; Ibrahim, 2016; Williams, Beardmore and Joinson, 2017; Workman, 2008),
7. Social Proof (Williams, Beardmore and Joinson, 2017; Workman, 2008),
8. Likability and Similarity (Luo et al, 2011; Williams, Beardmore and Joinson, 2017; Lederman, 2012; Workman, 2008),
9. Reciprocation (Ibrahim, 2016; Williams, Beardmore and Joinson, 2017; Workman, 2008),
10. Curiosity or Excitement (Kirlappos and Sasse, 2012; Ibrahim, 2016; Workman, 2008),
11. Commitment and Consistency (Williams, Beardmore and Joinson, 2017; Workman, 2008),
12. Overloading (Vishwanath et al, 2011),
13. Diffusion of Responsibility (Luo et al, 2011; Laribee et al, 2006),
14. Show-off e.g. heroism (Ibrahim 2016),
15. Convenience (Adams, 2012),
16. Interpersonal Relationships (Workman, 2008).

Next, the proposed conceptual framework and the instructional design for this research are discussed.

Chapter 3 Situational Awareness Model for Phishing (SAMFP)

3.1 Introduction

This chapter discusses the theoretical background underpinning the proposed conceptual framework called Situational Awareness Model for Phishing (SAMFP) which was developed during this research to help better understand the behavioural factors related to Phishing and improve awareness about them. SAMFP is the approach to answer the research questions and fulfil the research goals. According to the findings from literature (Proctor, Schultz and Vu, 2009; Anttila et al, 2007) and others, and as a practitioner in the field of Information Security, awareness has always been a concern since huge amounts of efforts put forth to protect information from cyber-attacks could be rendered useless if effective awareness outcomes are not achieved. These ineffective awareness results are either caused by lack of awareness or by providing awareness in an ineffective manner and due to these two reasons, the online learning environment or the workplace becomes exposed to the risk of Phishing despite all the technical controls in place. Mostly, awareness is provided using a one way communication in which the presenter delivers a speech or a presentation using highly technical jargon that not all attendees understand. In addition, most of these presentations talk about the visual symptoms of Phishing such as grammatical mistakes and typos ignoring the root cause manifested by the human behavioural factors exploited in Phishing attacks. As a result, the presentation or speech finishes with the audience not being able to clearly understand or tie what they have just learned to their living world in a way that makes sense to them so that cyber-attacks such as Phishing and its impact on information is correctly realized and assessed and that relevant mitigation controls such as awareness are appreciated. For this reason, it has been decided to make this attempt to better understand the reasons why such awareness efforts are ineffective and how to improve the learning outcomes of Phishing related awareness to provide a better human-based mitigating control against Phishing. Therefore this research's first question is two parts: 1) to identify the human behavioural factors that are vulnerable to exploitation in Phishing attacks targeting online learners and understand how they are exploited in Phishing attacks. 2) Explore and measure the level of awareness about such behavioural factors in a sample of online learners and find out conceptual ways to improve their awareness levels and test their effectiveness.

In order to answer these questions, a thorough literature review has been conducted to identify and analyse the human behavioural factors used in Phishing attacks with relation to the online learners' awareness needs. Furthermore, the findings from the literature review and the analysis of the human behavioural factors have been used to design the required awareness materials that address the identifiable online learners' needs. Not only that, but also helped in the design and development of the conceptual framework and the methodology proposed to develop the awareness program and deliver it. Thus, the research process consists of three elements:

- Building upon the literature review foundation (Ellis and Levy, 2008) which included the identification of the sixteen (16) human behavioural factors and based the design and development of the conceptual framework on Endsley's Situation Awareness Model (Endsley, 2015) and the pedagogical guidelines of Chen, Shaw and Yang (2006) to design and conduct the awareness program sessions and finally the Poepjes' (2012) Information Security Awareness and Capability Model (ISACM) for the assessment of the learning outcomes.
- Conducting the awareness program according to the proposed conceptual framework (SAMFP) to collect the required data through three (3) assessment tests.
- Evaluating the collected data through statistical quantitative analysis of the three (3) assessment tests' scores.

This research is considered a Design Based Research (DBR) since it meets the three elements of the Design Based Research emphasized by Andriessen (2007). These three elements are:

- The researcher is the designer of a solution to the problem which is in this research the instructional design of a full awareness program with all its resources and materials including its assessment tests.

- The solution is explicitly built upon the redesign of already known conceptual frameworks which are in this research derived from the integration of Endsley’s Situation Awareness model (2015), the proposed guidelines for delivering and communicating awareness by (Chen, Shaw and Yang, 2006) and finally the Poepjes’ (2012) Information Security Awareness and Capability Model (ISACM) which is interwoven in the design as part of the awareness measurement tool.
- The designed solution is put to action to test the validity of the solution and measure its outcomes which is in this research fulfilled by the researcher orchestrating and conducting the designed awareness program and the participants being at the centre of the action while the researcher simultaneously collects the required data for assessing and measuring the learning outcomes through conducting three assessment tests and quantitatively analysing their results.

The Endsley (2015) Situation Awareness Model, guidelines of Chen, Shaw and Yang (2006) and Poepjes’ (2012) Information Security Awareness and Capability Model (ISACM) were adopted through the literature review of behavioural factors for their sustainability and dynamicity to underpin the design of this research’s conceptual framework to address the identified needs for enhancing online learners’ awareness about human behavioural factors used in Phishing attacks.

These three components have been integrated into a new model that is iterative in nature and capable of providing successive interactive awareness sessions based on the assessed learning outcomes after each awareness session given until the highest level of awareness is achieved and the marginal lack-of-awareness risk is reduced.

Finally, as a researcher, having undertaken awareness materials design, delivery and analysis has given the researcher an excellent opportunity to observe closely the learning outcomes as they develop, which was a fruitful experience. At the same time, this has put a responsibility on the researcher to take extra care to record events as they happen minimising influence on participants.

In this chapter, the theoretical background underpinning the proposed conceptual awareness model used to raise awareness about Phishing behavioural factors is discussed. In addition, other aspects of this research design will also be covered in the discussion.

3.2 Theoretical Background

It is very critical for a successful information security program to be accompanied by a conceptual model that effectively addresses awareness needs of stake holders in order to help them thoroughly understand the information security risks and their mitigating controls. Addressing awareness requires good understanding of how humans acquire and manage awareness (Poepjes 2012). Based on the literature review conducted, out of many awareness theories such as OODA loop (observe, orient, decide and act) (Brehmer, 2005) and PDCA cycle (plan, do, check and act) (Sokovic, Pavletic, and Pipan, 2010), Situation Awareness (SA) (Endsley, 2015) has prevailed in the area of awareness raising (Poepjes 2012). Situation Awareness (SA) is the perception of what is going on in a situation where the perception is governed by the time and space of the current situation (Endsley, 2015). Similarly, Poepjes (2012) also described Endsley's Situation Awareness as "being aware of information or cues in your environment, and then determining what might happen next". Situation Awareness has several theoretical models such as Situated SA, Distributed SA and other current models such as Endsley 1995 SA Model (Endsley, 2015). Endsley 1995 SA Model is meant to enhance individual and team awareness in fields like Aviation, Air Traffic Controlling, Power Plants and ship navigation (Endsley, 2015). Likewise here, SA is going to be used to dynamically assess participant awareness about Phishing and enhance their understanding of Phishing behavioural factors by the researcher developing and implementing a conceptual awareness model engaging participants in interactive awareness group discussions where the researcher is the teacher.

3.3 Situation Awareness Levels

Situation Awareness has 3 hierarchically ascending levels namely (1) Perception, (2) Comprehension and (3) Projection. Wright, Taekman and Endsley (2004) in their simulated medical environment study and Kalliniatis et al (2017) in their efforts to unify Endsley Situation Awareness (Endsley, 2015) and Distributed Situation Awareness (Statnton et al, 2006) described these Situation Awareness Levels as:

- Perception (Level 1): The baseline in understanding the status, attributes and dynamics of relevant elements in the environment or as the "ground truth" as Brynielsson and Varga (2016) explained it in their Cyber Situational Awareness Testing.
- Comprehension (Level 2): The level of understanding whereby a synthesis of disjointed Level 1 SA elements is achieved through the processes of pattern recognition, interpretation and evaluation.
- Projection (Level 3): The highest level of Situation Awareness where participants have the ability to project the future actions of the elements in the environment.

Endsley's SA Model (Endsley, 2015) is dynamic in that it dynamically addresses awareness gaps and maintains sustainability. Thus, its link to time and space becomes viably capable of providing continuity and sustainability of awareness in a constantly changing field like information security. In other words, Endsley's SA is a cyclic process that the more iterations of awareness sessions provided, the higher levels of awareness achieved e.g. the Projection level. Therefore, Endsley's Situation Awareness model (Endsley, 2015) is adopted as a suitable foundation for this research's proposed conceptual framework that is integrated with a set of guidelines recommended by Chen, Shaw and Yang (2006) to raise awareness about Phishing behavioural factors for users of online learning environments.

3.4 Proposed Conceptual Framework for Raising Awareness about Phishing Behavioural Factors (SAMFP)

In order to answer the Research question 2, a conceptual framework is proposed based on the integration of three components namely, Endsely's Situation Awareness model (2015), a set of pedagogical guidelines recommended by Chen, Shaw and Yang (2006) and Poepjes' (2012) Information Security Awareness and Capability Model (ISACM) to help raise and quantitatively measure awareness in online learners about the behavioural factors underlying Phishing attacks. Thus, this proposed framework is called "Situation Awareness Model for Phishing" (SAMFP). Figure 1 shows the proposed conceptual framework model SAMFP.

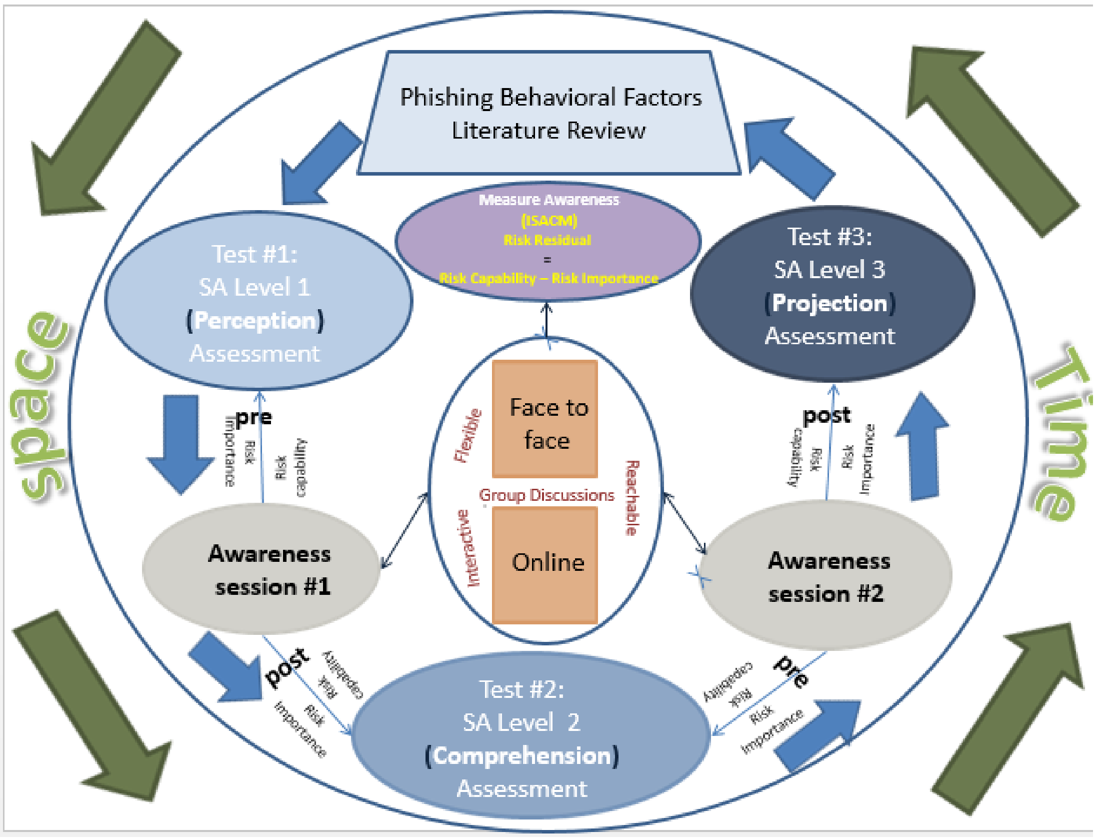

Phishing Behavioral Factors Literature Review
Measure Awareness (ISACM)
Risk Residual = Risk Capability – Risk Importance
Test #1: SA Level 1 (Perception) Assessment
Test #3: SA Level 3 (Projection) Assessment
space
Time
pre
Importance
Risk
Risk capability
post
capability
Risk
Risk
Importance
Flexible
Face to face
Group Discussions
Reachable
Interactive
Online
Awareness session #1
Awareness session #2
post
capability
Risk
Risk
Importance
pre
capability
Risk
Risk
Importance
Test #2: SA Level 2 (Comprehension) Assessment

Figure 1 - Proposed Conceptual Framework based on Endsley's Situation Awareness

Endsley's Situation Awareness levels (Perception, Comprehension and Projection) are woven into this model in order to achieve the objective of this research that is to enhance participant awareness levels about Phishing attacks and reach the Projection level. SAMFP model incorporates the following pedagogical guidelines (Chen, Shaw and Yang, 2006) into its fabric to provide an effective multi-faceted interactive awareness delivery methods in order for participants to achieve better awareness levels:

- Two-way communication as opposed to one-way whereby users can interactively participate.
- Create measureable targets to assess awareness program outcomes before and after.
- Flexible Awareness programs that can be modified as needed or as informed by the assessment outcomes.
- Reachable programs by introducing diverse methods of communication such as e-mail, wikis, blogs, discussion forums, online surveys and Face-to-Face presentations.

Hence, Online and Face-to-Face group discussions are utilized to enable interactive two-way communication to encourage collaborative discussions among participants during the delivery of awareness sessions. The categorization of online learners into the two groups namely the Online Group and the Face-to-Face group is based on the awareness delivery method. Thus, those participants who attended the awareness sessions via online means are categorized as the Online Group, while the participants attending the sessions in a classroom setup are called the Face-to-Face group. This categorization which is also the base of the scores analysis after the implementation of the SAMPF model aimed to explore and analyse whether the awareness delivery method being online or Face-to-Face could have impacted the participants' learning outcomes as a dynamic variable of the space. However, the results of the analysis showed non-significant implications of such an impact in the tests' scores despite the fact that the Online Group participants outperformed the Face-to-Face group participants. Yet, this outperformance of the Online group was found to be due to other significant factors such as the age, education level and the number of years using online learning as explained thoroughly in the Results Discussion Chapter.

The SAMFP's six phase process is designed to be flexible and reachable to accommodate participants' needs e.g. time zones, level of awareness etc. In the next section, the implementation of the SAMFP's six phase process is discussed in detail.

Finally, SAMFP model utilizes Poepjes' (2012) Information Security Awareness and Capability Model (ISACM) to quantitatively measure participant awareness improvements at the following 3 Endsley's Situation Awareness levels:

- Level 1, Perception: At this preliminary level of awareness, participants' perception about Phishing and its underlying behavioural factors is assessed.
- Level 2, Comprehension: At this level, participants' comprehension level (deeper understanding) of Phishing and its underlying behavioural factors is enhanced and assessed.
- Level 3, Projection: Participants are also assessed at the highest level of awareness for which they are targeted and thereby become able to predict Phishing attacks by recognizing behavioural factors in exploitable situations.

SAMFP utilizes the three attributes of ISACM (Poepjes, 2012) namely, Awareness Capability, Awareness Importance which are mapped to the gradual levels of Endsley's Situation Awareness (2015) to calculate the third attribute which is Awareness Risk Residuals based on participants' scores in the three assessment tests and the weighted difficulty levels for each test. Hence, the 3 Situation Awareness Levels are designed with a pre and post assessment tests surrounding each level. Detailed discussion of how SAMFP assigns weights to tests' difficulty levels and calculates scores, assesses and measures awareness improvements can be found in Chapters 4 and 5, sections 4.13, 4.14, 4.15 and 5.19.

The proposed conceptual framework (SAMFP) is time and space based which means that time and space are two dynamic variables that have direct impact on the learning outcomes. The effect of the time variable can be manifested by the cyclic nature of the SAMFP model which facilitates iterated awareness sessions, while the effect of the space variable is manifested by the flexibility of the SAMFP model to facilitate awareness sessions using face-to-face and online settings. Hence, SAMFP uses these two variables to maintain continuity and sustainability for information security awareness provided to participants across time and space to ensure better learning outcomes and measured gradual awareness improvements until the highest level of awareness ‘Projection’ is achieved.

While the proposed conceptual framework SAMFP is based on a combination of learning theories and guidelines from literature. It is distinctive by the following features:

- It provides a collective set of features that may be found in one tool and missing in another.
- It adopted the best features from each of the underlying learning models and at the same time complemented them by adding the other features from the other models. For example, SAMFP adopted the situation awareness levels and their space and time attributes from Endsley’s Situation Awareness Model (Endsley, 2015) and enhanced it by devising a novel measurement technique to measure and track awareness progress and awareness risk residuals based on the ISACM (Poepjes, 2012).
- In addition, the SAMFP model ensures continuous assessment of the participants’ level of awareness about each behavioural factor individually after each test which also can help in directing the focus for the awareness materials development in the right direction of participants’ learning needs in the next awareness sessions to follow in the SAMFP cycle.
- Furthermore, the Endsley’s Situation Awareness model and even the ISACM (Poepjes, 2012) did not define a process for developing and conducting awareness, while the proposed SAMFP has provided pedagogical guidelines to inform the process of awareness program development and delivery.

- In addition, the SAMFP model ensures continuous assessment of the participants' level of awareness about each behavioural factor individually after each test which also can help in directing the focus for the awareness materials development in the right direction of partivcipants' learning needs in the awareness sessions to follow in the SAMFP cycle.
- Finally, the SAMFP also ensures maintainability of the identified behavioural factors by injecting a periodic review as part of its cycle of emerging literature about behavioural factors to ensure that SAMFP will always remain up-to-date in terms of the knowledge it provides.

According to Figure 1, the SAMFP model has been implemented in 6 steps which are discussed next.

3.5 SAMFP Implementation

In order to implement the SAMFP model, the following 6 steps have been implemented:

1. Conduct Literature Review to Identify Phishing Exploitable Behavioural Factors: Conducted a literature review to explore existing literature such as security breach reports and analytical studies on Phishing using behavioural factors to lure unsuspecting victims. The objective of the literature review was to identify and understand the potentially vulnerable behavioural factors that could be exploited in Phishing attacks. The outcome from this literature review is 16 potentially behavioural factors forming the basis for developing the awareness materials for the awareness sessions, group discussions and participant awareness level assessment tests.

2. Develop and Conduct SA Level (1) Perception Assessment Test - the 1st Test: This assessment test is the 1st Test which is one of 3 online surveys (tests) that serves three purposes. First, it provides an introduction about the research in general and the assessment test in particular and obtain the participant's consent to proceed with his/her participation in the research. Second, it gathers participants' demographical information such as age group and education level. Third is the preliminary assessment test which makes the main content of the 1st Test. The 1st Test poses as (1) the initial assessment test in the data collection phase by which participants' preliminary understanding (SA Perception Level) about Phishing is assessed and (2) as the pre-awareness assessment test before participants attend the first awareness session and take the post assessment test afterwards (2nd Test). The objective of the pre-awareness test is to quantitatively assess the preliminary awareness level (Perception Level in the Conceptual Framework) of participants about Phishing in general before attending any awareness sessions. The 1st Test consists of 20 fictitious scenarios whereof 16 scenarios are potentially Phishing scenarios demonstrating the exploitation of one of the 16 Phishing behavioural factors identified in the literature review. Participants are assessed on whether they are able to identify, based on their preliminary understanding, if the scenario represents a Phishing or Non-Phishing situation. If a participant chooses 'Phishing' as the answer, he/she is required to explain in their own words the reasons for choosing such an answer. This is to assess how relatively close the participants' own rationale to identify phishing is to the actual behavioural factor exploited in the Phishing scenario. Data from this assessment test will be analysed to answer Research Question 1.

3. Conduct the 1st Awareness Session as Online and Face-to-Face Group Discussions: In order to help participants gain better understanding of how the 16 behavioural factors are exploited in the Phishing scenarios after their initial assessment in the 1st Test, they are scheduled to attend their first awareness session. The 1st awareness session is designed as an interactive group discussion to encourage participants to share their own ideas and experiences about Phishing and openly discuss the risks and dynamics of Phishing attacks reflecting on the 20 fictitious scenarios from the 1st Test and the 16 behavioural factors employed in each scenario. As key to the SAMFP implementation, this awareness session is a necessary step helping participants move from the 'Perception' Level to the 'Comprehension' level where they can have more of an in-depth understanding about the 16 behavioural factors and how they are exploited to deceive victims in the Phishing scenarios. The awareness session is set up either as an online group discussion using Google Hangouts or as a Face-to-Face classroom based meeting. This is to accommodate the variability of time and space as far as the SAMFP model is concerned. The setup of the online group discussion accommodates 5 to 8 participants each for those who are not able to attend in person, while the classroom based setup accommodates participants in a Face-to-Face setting since they can attend in person. In both setups, participants are engaged in interactive discussions and role play acting where a participant is asked to play the role of the attacker while another participant plays the role of the victim and demonstrate different ways of exploiting the 16 behavioural factors similar to those demonstrated in the 1st Test's scenarios. This approach is helpful for participants to develop an in-depth understanding about the 16 Phishing behavioural factors which helps them gradually achieve the next level of awareness namely Comprehension. In addition, a thorough review followed by a discussion of the 20 fictitious scenarios from the 1st Test was conducted to allow participants to contrast their answers before and after attending the awareness session.

4. Develop and Conduct SA Level (2) Comprehension Assessment Test - 2nd Test: This is the 2nd assessment test which aims to verify whether the participants' awareness about Phishing has reached the Comprehension level. The 2nd Test is considered the post-awareness assessment test after the first awareness session and at the same time the pre-awareness assessment test before the second awareness session. The 2nd Test is based upon the same 16 behavioural factors employed in the 1st Test's scenarios. However, the 16 behavioural factors are re-arranged differently and merged into 7 complex scenarios such that 2 or 3 behavioural factors are combined into one scenario. This is to add a layer of difficulty to the assessment in the 2nd Test in proportion to the 'Comprehension' level (Level 2) which is a higher level on the SA based conceptual framework. Thus, each of the 7 Phishing scenarios on the 2nd Test demonstrates the exploitation of 2 or 3 behavioural factors. Unlike the 1st Test, on the 2nd Test, participants are asked to choose the correct combination of behavioural factors that are exploited in the scenario from a pre-defined list of options. In contrast to the 1st Test where participants were only asked to identify if a scenario was Phishing or not, on the 2nd Test, they are asked to identify the behavioural factors exploited in each scenario by using the knowledge gained from attending the 1st awareness session and by demonstrating their in-depth understanding. This is to assess whether they achieved the 'Comprehension' level of Situational Awareness. However, unlike the 1st Test, participants on the 2nd Test are not required to enter any additional information to explain why they chose a certain answer.

5. Conduct 2nd Awareness Session as Online and Face-to-Face Group Discussions: Similar to the 1st Awareness Session, the 2nd Awareness Session is attended by participants after completing the 2nd Test and before taking the 3rd assessment Test. The objective of the 2nd Awareness Session is to help participants improve their understanding and knowledge about the 16 behavioural factors even better and take it to the next level of ‘Projection’ on the Situation Awareness model. Achieving the ‘Projection’ level of situation awareness means that participants will not only be able to recognize the exploited behavioural factors in Phishing scenarios, but also will be able to predict how the 16 behavioural factors will be used in future Phishing attacks. The 2nd Awareness Session is designed and set up similarly to the 1st Awareness Session except that the participants will be discussing the 7 scenarios on the 2nd Test instead.

6. Develop and Conduct SA Level (3) Projection Assessment Test - 3rd Test: This is the third and final assessment test. The 3rd Test concludes participants' participation and demarcates the end of the data collection phase. The 3rd Test is also the post-awareness assessment test taken after the 2nd Awareness Session. Similar to the 2nd Test, the 3rd Test consists of only 7 scenarios with a pre-defined list of answers combining 2 or 3 of the 16 behavioural factors which participants will choose from in order to answer the questions on the test. Unlike the 2nd Test's scenarios, on the 3rd Test, the scenarios are composed to potentially exploit a vulnerable situation where participants are asked to select the most suitable combination of behavioural factors to be exploited in the vulnerable situation of the scenario. This is to add a layer of difficulty to the 3rd Assessment Test in proportion to the ‘Projection’ level (Level 3) of the SA in the SAMFP model. In other words, the 7 scenarios are set up to assess whether participants had achieved the ‘Projection’ level of awareness after attending the 2nd Awareness Session. This is assessed by participants demonstrating their ability to predict the behavioural factors most suitable for exploiting a vulnerable situation compared to their ability to only identify behavioural factors already exploited in a situation as those in the 2nd Test's scenarios.

3.6 Research Methods

Quantitative Analysis (Cohen, Manion and Morrison, 2011) is used as the primary method to analyse the collected data from the three assessment tests. Quantitative Analysis is used to analyse the scores from the three pre and post-awareness assessment tests in order to measure the improvements gained by the participants before and after attending the 1st and 2nd awareness sessions respectively. This is achieved by comparing the scores between the 3 assessment tests using statistical variance analysis tests such as t-tests and Analysis of Variance (ANOVA) tests.

The secondary method of analysis is the qualitative thematic analysis using the "Essentialist" method with which the meaning of the participants' response is captured and linked to the overall context of the research (Braun and Clarke, 2006), which is the Phishing behavioural factors. As such, the meaning of participants' comments from the 1st Test are categorized under general themes related to the Phishing context. These themes were coded from participants' comments using the 'Inductive' approach (Braun and Clarke, 2006). These comments were entered by participants to explain why they thought a scenario on the 1st Test was Phishing. Then those general themes were quantitatively analysed and correlated with the participants' scores to verify whether the participants had answered the 1st Test's questions based on correct perception of Phishing behavioural factors (perception level) or based on having a different interpretation of the scenario in the question. The general themes and the analysis of the participants' comments will be further discussed in Chapter 5.

Finally, all the results from the quantitative analysis tests are correlated in search of potential relationships that would support the final research findings. Next, Situation Awareness measurement techniques are discussed.

3.7 Situation Awareness Measurement

There are several measurement techniques in the literature proposed for the assessment and measurement of situation awareness. However, only a few are suggested by Endsley for measuring the Situation Awareness levels. These are the Situation Awareness Global Assessment Technique (SAGAT) and the Quantitative Analysis of Situational Awareness (QUASA) technique (Brynielsson and Varga, 2016). The SAGAT technique relies on probing the participants' understanding levels by intermittently asking them questions crafted to assess each level of situation awareness (Perception, Comprehension and Projection). These questions are geared to investigate carefully chosen parameters in a dynamic situation at random points of time (Brynielsson and Varga, 2016). On the other hand, the QUASA technique is a combined self-rating and probing technique by which participants are asked 'True/False' or 'Yes/No' questions to confirm their understanding (perception level), thus the probe. Then, they are asked to explain and state why they had chosen a certain answer and on what basis they were guided by the scenario towards choosing that answer, hence the self-rating (Brynielsson and Varga, 2016). Brynielsson and Varga (2016) suggested at the end of their discussion of situation awareness measurements, that there is no method that is perfect for one discipline. Thus, they suggest that a customized measurement technique is devised to suit a domain such as the cyber security domain under assessment. Therefore, SAGAT and QUASA techniques were both capitalized upon in this research to measure participant awareness levels about Phishing attacks and the behavioural factors behind them. Accordingly, on the 1st Test, participants were asked to confirm whether a given scenario was a Phishing scenario by answering a 'True/False' (e.g. 'Phishing/Non-Phishing) question and then support their chosen answer in the case of confirmation (e.g. if it was a Phishing scenario) by stating their reasons for choosing that answer. On the other hand, on the 2nd and 3rd Tests, participants are only asked to choose from a drop-down list the correct answer.

Based on Endsley's Situation Awareness, Poepjes (2012) introduced the Information Security Awareness and Capability Model (ISACM) which capitalizes on the integration of three attributes of situational awareness to measure the achievements gained in the three awareness levels. These attributes are (1) the Awareness Importance (Awareness criticality level based on given weights) of the controls required to prevent a risk such as Phishing from occurring, (2) the Awareness Capability which is the participants' knowledge or existing level of awareness about the risk and the required control to prevent it and finally (3) the Awareness Risk which is the difference between the Awareness Importance and the Awareness Capability. ISACM (Poepjes, 2012) is adopted in this research to measure the level of the residual awareness risk after the participant awareness levels are assessed against Endsley SA levels using the weights assigned to each level. Thus, when the different levels of Situation Awareness are measured, normally pre and post assessment tests are conducted and their respective scores are compared. To ensure fair comparison of scores between pre and post-tests, weights are used to factor in the varying difficulty levels in each test (Kalliniatis et al, 2017; Poepjes, 2012). The formulated Situation Awareness Weighted Network (SAWN) model by (Kalliniatis et al, 2017) adds incremental weights to the Situation Awareness levels. Respondents to SAWN surveys found them very challenging due to the subtle incremental progression (Kalliniatis et al, 2017).

Similarly, the three assessment tests in this research are designed to progress incrementally from Perception through Comprehension to Projection with incremental weights to match up to the three situation awareness levels manifested by each test. The responses on each test are scored using a 7 point scale where 1 represents the lowest and 7 the highest. According to the SAWN model, the questions on the 3 assessment tests are constructed based on hypothetical scenarios that encapsulate different fictitious phishing attacks against which participant awareness is assessed for the three Situation Awareness Levels of the underlying (proposed) conceptual framework. Details of the implementation of the situation awareness measurement technique adopted in this research will be discussed in Chapter 5. Next, the limitations of the proposed conceptual framework are discussed.

3.8 Limitations of the Proposed Conceptual Framework

Endsley's Situation Awareness Model (2015) on which the proposed conceptual framework SAMFP is based, is considered effective in raising awareness (Endsley, 2015). However, within the scope of this study, the following limitations have been identified:

- The awareness' improvements measured during the implementation of the SAMFP model for the 100 participants were at varying levels; meaning that the improvement in awareness is achieved gradually and relative to the changing time and space variables of the SAMFP model. Hence, the SAMFP model was effective in helping participants make quantifiable progress towards achieving better levels of awareness in relation to what they were initially assessed for; with every iteration of awareness conducted. The limitation to this implementation of the SAMFP model lies in the constraint of the timeframe allowed for the research which restricted the awareness/assessment iterations to two iterations only. However, had there been more time allowed, more iterations could have been conducted.
- Another limitation or rather a challenge was to conduct the online awareness sessions and group discussions for participants at different time zones and behind virtual boundaries, which may have not helped fulfil the intimacy of Face-to-Face discussions. However and aside from that, the nature of the study targeting online students, makes this limitation an opportunity for testing the effectiveness of the proposed conceptual framework when used in online settings as opposed to Face-to-Face settings. To reduce the impact of this limitation, awareness was conducted to manageable groups of 10 so that participants from different time zones can be accommodated in a more flexible manner.
- Social detachment while conducting the online sessions to the Online Group participants was sensed, although not statistically shown to be significant, during the implementation of the SAMFP as compared to the Face-to-Face sessions.

Chapter 4 Data Collection and Score Standardization

4.1 Introduction

Phishing cannot be prevented by depending on technical controls only (Proctor, Schultz and Vu, 2009; Anttila et al, 2007). Effective Information Security Awareness is a key mitigating control against Phishing (Chen, Shaw and Yang, 2006) and to many other malicious attacks that come after it. However, according to Proctor et al (2009), Anttila et al (2007) and many others, the way information security is currently provided does not enable users to acquire in-depth understanding of how Phishing attacks are set up to apply this knowledge in the real world to protect themselves from Phishing. This is because many cyber awareness programs tend to overlook human behavioural factors as a root cause of exploitation in Phishing and focus mainly on teaching their audiences how to identity Phishing by superficial signs, such as hidden URLs, or fabricated logos (Kirlappos and Sasse, 2012). These superficial signs today can be sleekly overcome by the use of advanced technology and sophisticated tools deployed by attackers to craft solid-proof phishing e-mails. Therefore, this research aims to better understand these behavioural factors and their exploitation in Phishing in order to improve information security awareness training for online learners. Hence, the research questions are to 1) identify the human behavioural factors that are vulnerable to exploitation in Phishing attacks targeting online learners and understand how they are exploited in Phishing attacks, 2) Explore and measure the level of awareness about these behavioural factors in a sample of online learners and how susceptible they are to Phishing attacks, and research, and test the effectiveness of, a conceptual framework that can be used in the design of training to improve the awareness levels of learners. In order to answer the research questions, a thorough literature review has been conducted to identify and analyse the human behavioural factors used in Phishing attacks with relation to the online learners' awareness needs. Furthermore, the literature review and the analysis of the human behavioural factors have been used to design and implement a conceptual framework called 'Security Awareness Model for Phishing' (SAMFP) based on the integration of Endsley's Situation Awareness model (Endsley, 2015), the awareness delivery guidelines by (Chen, Shaw and Yang, 2006) and the Poepjes' (2012) Information Security Awareness and Capability Model (ISACM). The conceptual framework has been used to address the identified needs of online learners to improve their awareness about Phishing behavioural factors and assess the effectiveness of the learning outcomes after the implementation of the SAMFP model; which provided answers to the second research question. The implementation of the SAMFP model involved the researcher in designing and developing awareness materials, assessment test content and measures and finally conducting the training.

This chapter discusses participant recruitment, ethical issues, participation issues and the researcher's involvement and role in the research, followed by a discussion of the development of data collection strategies, tools and methods and finally the assessment tests' scores standardization and tests' difficulty levels weighting.

4.2 Participant Recruitment

Participants for this project were online students who were recruited based on self-selection (Saunders, 2012). 750 participants were recruited, but only 100 participants successfully completed all 5 steps required. Many of the recruited participants gave up participation after completing the 1st Assessment Test. It felt great to have achieved 100 complete participations after all; over a period of 8 months only. According to Delice (2010), the number of participants required for quantitative analysis to produce meaningful results with acceptable level of confidence (5%) should be in the range between 30 to 500 participants. Therefore, a target of 100 participants was set and when this was achieved the data collection process stopped.

4.3 Participants Selection Strategy

Despite the argument in literature about using convenience or non-probability versus purposive sampling in Social qualitative research being subject to the risk of selection bias (Etikan, 2016;; Coyne, 1997), it is considered an acceptable alternative sampling approach when used with compensating precautions.(Magnani, Sabin, Saidel and Heckathorn, 2005). Therefore, considering the fact that this research is a quantitative study that focuses on participants being online learners only regardless of any other traits; combined with the hardship associated with participants recruitment given the timeframe allotted, this research used convenience sampling to select the minimum required number of participants which is a 100 who should complete the 5 required steps of participation in order for their data to be considered for analysis.

Therefore, an invitation for participation with all the details of the 5 steps was sent to several online student communities including the local online university of Taybah based in Saudi Arabia', multi-nationality academic research collaborative network e-mailing lists such as the UK based JISCMail,lists, multi-nationality LinkedIn accounts, multi-nationality Twitter groups such as the UK based Twitter account of @PhDForrum group which has members from all over the world, the US and UK based Saudi Students' Twitter groups accounts and finally the Online Learning department of a local Saudi based company.

In line with the convenience sampling approach, the invitations were sent to reach out for any potentially candidate source of recruitment for online learners. As a result, 750 participants were recruited; however, only 100 participants completed all the required steps of participation. These 100 participants are located based on the convenience sampling approach, mainly in Saudi Arabia, the UK and USA.

It is worth highlighting that the only requirement for a participant to be selected to participate in this study is to be an online learner regardless of location or culture. Therefore, the distribution of participants in terms of location and culture was not part of the selection strategy or the participation criteria. As a result, there was no ground for any solid implications to be elicited in the data analysis that could support any significant conclusions based on the existing sample distribution. Furthermore, the distribution of the sampled 100 participants in terms of other demographics such as the job and education field did not reveal any significance in the study due to inconsistent distribution of participants in terms of these two demographics. While this is a limitation of the study, this could also be a potential opportunity to extend this in future research studying the effect of a particular sampling distribution that is based on specific demographics such as culture, nationality, location, education major or career on online learners' susceptibility to Phishing attacks.

Based on the awareness delivery method used, 40 participants out of the 100 were categorized as the Online participants group since they were given the awareness sessions via an online medium, while the remaining 60 participants who are also online learners were categorized as the ‘Face-to-Face participants group since they were given awareness using a classroom based setup.

The process of participant recruitment took many cycles of communication back and forth with interested members from the invited groups to answer any questions and clarify any concerns raised. The process of recruitment and participant selection was ongoing iteratively even during and after the data collection phase had started. This was due to some of the participants withdrawing from participation or not completing their participation bringing down the number of completed participations below the required minimum number. After 8 months of continuous coordination, the required number of completed participants successfully reached 100.

4.4 Participants' Demographics

Participants were recruited on the condition that they were or had been part of an online learning course. 100 participants were finally considered for the research activities. These 100 participants were divided into two groups with 40 participants in the Online Group and 60 participants in the Face-to-Face group. The majority of participants were males and aged between 20-51 years old. However, the majority of them (74%) aged between 20-25 years old. The participants' education levels varied between High School and PhD with the majorly (57%) being in the High School level and 32% with bachelor degrees while the rest have higher educational degrees such as Masters and PhD. The participants also came from a wide variety of work and study fields and backgrounds ranging from trainees in Oil and Gas to part-time students in fields such as Human Resource, IT and Engineering. Participants also had different responsibilities at their workplaces ranging from leadership and management positions to trainees and technicians. Participants' demographics will be analysed and discussed in depth in Chapter 5.

4.5 Ethical Issues

This awareness program is a preventive measure against Phishing attacks, however, the improper knowledge resulting from not fully completing the program could potentially increase the sensitivity of participants to Phishing attacks. On the other hand, gaining the knowledge with lack of accountability could also introduce a risk of abuse or malicious use of the knowledge. Therefore, as a precautionary measure, participants should be made aware of their accountability and need to ensure due diligence for proper use of the knowledge shared and that they should always adhere to secure practices. This study was given ethical approval by Lancaster University FASS-LUMS Ethics Committee, see appendix D.

Other than the above, no major ethical issues emerged during the conduct of the research because ethics related matters had been considered in the design of the research from the very beginning. Preventive measures had been established to protect and handle ethical considerations such as participants' information confidentiality which was taken care of in the participation consent obtained as part of the 1st Test (See appendix C). The 1st Test was designed to include the participation consent form by which ethical considerations and conditions were explained to participants as participants were free to accept or reject the consent statement. Participants were assured that their information would be classified as confidential information and anonymized, and that it would only be used for data analysis in this research. In addition, participants could choose to withdraw from the research at any time before data analysis commenced. Participants were also assured that their data would be securely handled and retained anonymously in secure media storage. Videotaped footage of the online group discussions during the awareness sessions and the classroom based sessions were also stored in secure storage media as participants were assured that these videos would only be used in the research data analysis and would not be disclosed outside that scope. All information of participants who failed to complete the 5 steps, but agreed to the consent was dropped completely from the research, however securely retained.

4.6 Researcher’s Role

Combining Action Research with a Design Based Research (Andriessen, 2007) to execute and enact the design, the researcher was heavily engaged in the design, development and implementation of the research action plan. The researcher has developed the awareness materials and assessment tests’ content and measurements, tutored awareness sessions online and at a classroom. These activities the research took on were part of implementing the conceptual framework designed for conducting this research. Finally, the researcher’s role was to conduct data analysis to verify and validate the learning outcomes from implementing the conceptual framework against the research questions and objectives. In this research type and environment, where the researcher is heavily engaged in conducting the research and particularly when the researcher had practiced tutoring online and in a classroom, the risk of bias (Elliot, 1999) is high as the researcher is the designer and the action taker at the same time (McKay and Marshall, 2001) knowing exactly what needs to be achieved. The researcher might unconsciously drive the research and the results thereof to where the results could meet the objectives of the research. Therefore, the researcher has always taken extreme care and constantly reminded himself to remain a diligent, and reflective, observer of personal and participants’ actions while simultaneously being a moderator for action. This has allowed the researcher to observe participants’ reactions and responses during awareness group discussions more closely with an eye on the researcher as well to avoid directing participants’ actions while recording observations as they are to ensure the reliability and validity of the data collected for analysis.

Moreover, the experience of tutoring online and in a classroom added a great value to the researcher's engagement by deepening the researcher's understanding of how knowledge was constructed through collaborative information sharing and exchanging individual experiences of participants as well as tutors in the research. Part of this added value was in understanding the differences between the online and classroom, or Face-to-Face, environments as far as tutoring and knowledge construction are concerned including the reason for Blended Learning emerging to compensate for such differences (Rovai and Jordan, 2004). An example of these differences observed was the loss of facial expressions and eye contact in the online virtual environments which might have caused detachment from the online group and impacted the learning outcomes and the effectiveness of the knowledge construction process.

4.7 Validity and Reliability

Validity and reliability are two important attributes of quantitative research (Golafshani, 2003; Winter, 2000) in which research instruments, content and results should demonstrate an acceptable level of validity namely instrument validity, content validity and external validity respectively (Winter, 2000).

Although Factor Analysis is a reliable method to assess validity in quantitative research, it was not used in this research since the sample size in this research is 100 participants is not large enough to meet the prerequisite of using Factor Analysis which is 10-15 participants per item in the questionnaire in order to produce meaningful results (Melainie, 2012). Although, Factor Analysis can be used to measure validity of binary datasets such as the True/False questions used in the dataset of this study (Goforth, 2015), yet it is more efficiently used with a dataset featuring scaled responses and continuous variables such as those using Likert Scales that measures for example, attitude, favouritism, pleasure (Melainie, 2012).

With regards to using Cronbach's alpha for assessing reliability, the reverse-phrased scores in the dataset being either 0 or 1 will produce negative covariance between the scores which will render lower and consequently incorrect Cronback's alpha (Melainie, 2012). For these reasons, qualitative validity and reliability methods are used.

Instrument validity is demonstrated by developing the tools used for collecting the data in this research using SurveyMonkey which is a reliable and trustworthy online platform for developing surveys and questionnaires. These tools were used as the means for conducting the three assessment tests through which the research data was collected. On the other hand, content validity was manifested by the content of the three assessment tests and the awareness materials which were developed in conformity with the findings from the literature review as far as the research objectives and questions were concerned. The content of the three assessment tests and the awareness materials was developed to be consistent with and reflective of the behavioural factors identified through the literature review in response to the identified online learners' awareness needs about exploiting human behavioural factors in Phishing attacks. Based on the adopted Situation Awareness Model of Endsley (2015), the content of the tests and awareness sessions demonstrated the exploitation of behavioural factors using fictitious scenarios similar to the real-world phishing incidents reported in security reports found in the literature. The similarity of the fictitious scenarios to the real-world attacks was meant to ensure the validity of the tests' content or in other words the content of the research instruments used.

Finally, the external validity of the collected data which lies in the generalizability of the research findings from the research sample to the full population (Merriam, 1995) of online learners was considered from the beginning by selecting a representative sample of the population. The sample was selected based on one criteria which was a participant must be an online learner or have spent some time in online learning. The number of participants selected was 100, which is considered to be within the acceptable range of 30 to 500 participants with (5%) of confidence to produce valid and reliable results that are generalizable (Delice 2010).

The reliability of the data is concerned with the consistency of the measurements used to assess and measure the data and the outcomes of data analysis in the research (Winter, 2000). In this research, the data collected from the three assessment tests are the participants' scores. Two measures were taken to ensure consistency of the measurements and thus the reliability of the data in this research (Burke, Fahn, Marsden, Bressman, Moskowitz, and Friedman, 1985). First, the scoring scales were standardized among the three tests to ensure equality and norm. Second, weight factors were assigned to each test respective to the test's incremental difficulty level that had been built into its content and design. These weights were factored in the calculation of the scores of each test to ensure fairness and consistency of score measurement among all three tests. The quality and credibility of the collected data depended so much on the design of the tests to encourage participants to answer the questions by posing user friendly designs.

4.8 Data Collection Strategy

Data collection started as soon as the literature review started as a preparatory step for further data gathering at a later stage. During the literature review, relevant materials on Phishing behavioural factors were consulted. The outcome of the literature review was the identification of 16 exploitable behavioural factors behind tricking victims in Phishing attacks. See Chapter 2 for details on these behavioural factors.

1. Temptation.
2. Urgency or Scarcity.
3. Over-confidence or Self-Consciousness.
4. Over Trust.
5. Authority.
6. Threatening, Fear or Anxiety.
7. Social Proof.
8. Likability and Similarity.
9. Reciprocation.
10. Curiosity or Excitement.
11. Commitment and Consistency.
12. Overloading.
13. Diffusion of Responsibility.
14. Show-off.

15. Convenience.

16. Interpersonal Relationships.

All the above-mentioned behavioural factors were used as the basis for developing the awareness course materials and the assessment tests used during the data collection phase which is explained next. As explained in Chapter 3, the data collection procedure involved carrying out 5 activities in addition to the initial literature review. These are:

1. Develop and conduct the 1st Test, including, collecting participants' consent and demographic information.
2. Develop the material for Awareness Session #1 and conduct it.
3. Develop and conduct the 2nd Test.
4. Develop the material for Awareness Session #2 and conduct it.
5. Develop and conduct the 3rd Test.

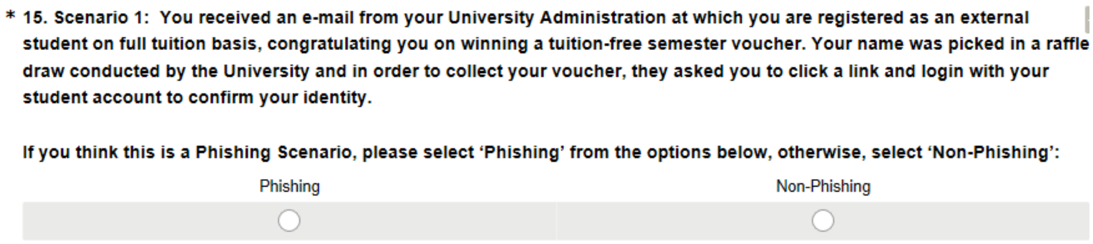
Five tools corresponding to each step of the data collection procedure were developed. Thus, three surveys were developed as the tools used to facilitate the 3 tests and two Power Point presentations contained the course material for conducting the two awareness sessions. A detailed discussion of the development of these 5 tools is next.

4.9 Data Collection Tools

First, three surveys were developed to facilitate collecting participants' consent to participate in the research, participants' demographic information (Categorical data about the participants such as years of experience being an online student, education level and age). This is in addition to conducting the three assessment tests to measure participant awareness and understanding about the 16 behavioural factors behind Phishing attacks identified during the literature review. SurveyMonkey (Waclawski, 2012) was used to develop the three surveys/tests which are explained in detail next.

SurveyMonkey Survey - the 1st Test: The 1st Test serves three purposes. First, it provides the means to get the participant's consent to participate in the research. Secondly, it gathers demographical information about the participants such as age, and education. Thirdly, it delivers the 1st assessment test which serves as the pre-awareness test taken by participants before attending the planned two Phishing awareness sessions. The objective of the 1st Test is to assess and measure the participants' preliminary level of understanding about Phishing. This preliminary level of understanding is called 'Perception' according to Endsley's Situation Awareness Model (Endsley, 2015) ;previously explained in Chapter 3.

The 1st Test comprises 20 questions. Each question is composed of an imaginary scenario manifesting one of the 16 Phishing behavioural factors as the Phishing bait to trick potential victims to fall prey to it. 16 of these 20 questions demonstrate a Phishing scenario, while 4 questions demonstrate non-Phishing normal scenarios inserted between the test's questions just to create the required testing effect. The 20 questions are Dichotomous questions (Cohen, Manion and Morrison, 2011) which require a nominal answer provided in a selectable answer list as 'Phishing' or 'Non-Phishing' (see Figures 2 and 3). Thus, each question requires the participants to carefully read the scenario, evaluate it and provide an answer by choosing one of the selectable answers describing the scenario as Phishing or Non-Phishing based on the participant's preliminary understanding of Phishing and evaluation of the scenario. As far as the 1st Test is concerned, the assessment of the participants' preliminary understanding of Phishing requires the participants to demonstrate their ability to distinguish a Phishing Scenario from a Non-Phishing scenario without being asked to identify the Phishing behavioural factors underlying each scenario at this preliminary stage. However, in the event they chose 'Phishing' as the answer, they were required to answer a follow-up open-ended question explaining in their own words the reasons for choosing 'Phishing' as the answer; or in other words, their own rationale for why they believed the scenario was Phishing. For example in a Phishing scenario, the attacker used temptation (Jagatic et al, 2007; Kirlappos and Sasse, 2012) as the bait to lure victims into the Phishing trap. Temptation is the underlying behavioural factor that was used as an incentive to manipulate victims and impact their logical decision making process so that they would fall into the trap (Thabtah and McCluskey, 2012). If the participant chose 'Phishing' as the answer, he or she was required to support their answer by stating the reasons for choosing that answer. In the Non-Phishing scenario, there are no implied behavioural factors involved that would hint Phishing to the participant. It is only there to test the participant's ability to distinguish Phishing from non-Phishing scenarios.

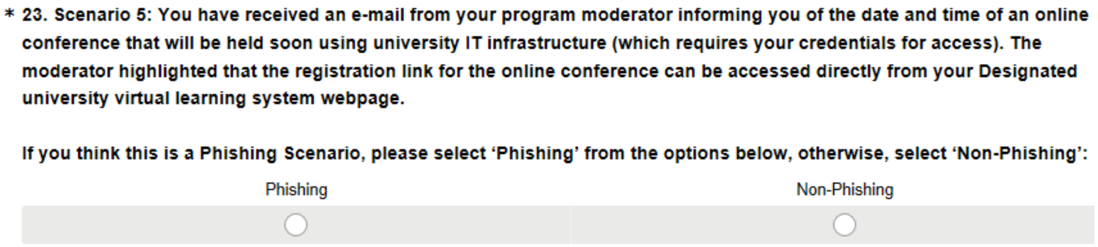


* 15. Scenario 1: You received an e-mail from your University Administration at which you are registered as an external student on full tuition basis, congratulating you on winning a tuition-free semester voucher. Your name was picked in a raffle draw conducted by the University and in order to collect your voucher, they asked you to click a link and login with your student account to confirm your identity.

If you think this is a Phishing Scenario, please select 'Phishing' from the options below, otherwise, select 'Non-Phishing':

Phishing	Non-Phishing
○	○

Figure 2 - Example of 1st Test's Scenarios



* 23. Scenario 5: You have received an e-mail from your program moderator informing you of the date and time of an online conference that will be held soon using university IT infrastructure (which requires your credentials for access). The moderator highlighted that the registration link for the online conference can be accessed directly from your Designated university virtual learning system webpage.

If you think this is a Phishing Scenario, please select 'Phishing' from the options below, otherwise, select 'Non-Phishing':

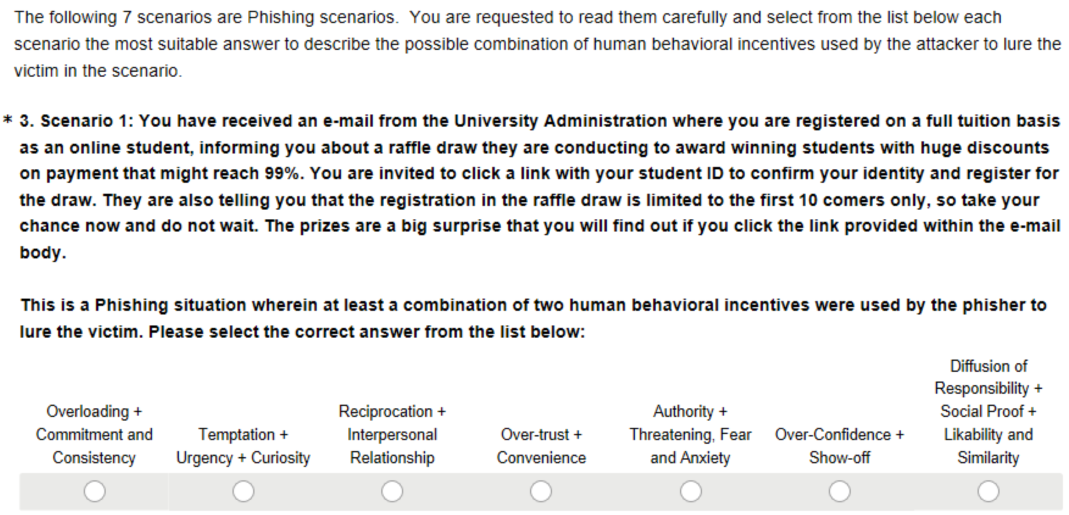
Phishing	Non-Phishing
○	○

Figure 3 -Example of 1st Test's Scenarios

SurveyMonkey Survey - the 2nd Test: As the 1st Test's objective was to provide for the pre-awareness assessment, the 2nd Test provides for the post-awareness assessment for participants after taking the first Phishing awareness session. The 2nd Test aims to assess whether there was a significant improvement in participants' understanding about Phishing after having attended the first awareness session by comparing their test scores before and after attending the awareness session; e.g. comparing their scores on the 1st Test (Pre-Awareness Assessment) with their scores on the 2nd Test (Post-Awareness Assessment). The participants in both tests were the same group; however their scores on the 1st test represented the control group (Fraenkel,, Wallen, and Hyun, 1993) to which their scores on the 2nd test were compared as the Experimental Group.

Unlike the 1st Test, the 2nd Test comprises 7 Phishing scenarios composed of combining two or three behavioural factors from the 16 factors into one scenario to bait the victim. Hence, the 2nd Test covers all the 16 behavioural factors using 7 scenarios only. However, the 2nd Test adds an extra level of difficulty to the scenarios to match up to the expectations of the second highest level of awareness according to the Situation Awareness Model (Endsley, 2015) namely 'Comprehension' against which participants in the 2nd Test are assessed. Yet, the 2nd Test maintains the similarity of content with the 1st Test. The extra level of difficulty in the 2nd Test is meant for testing a higher level of understanding in the participants expected to be gained after attending the first awareness session. In order to quantitatively measure the participant awareness improvements gained between the two tests, the participants' scores on each test are factored by a weight assigned to the level of difficulty manifested by the test corresponding to the level of awareness assessed respective to Endsley's Situation Awareness levels (Endsley, 2015) which is Perception in the 1st Test, Comprehension in the 2nd Test and Projection in the 3rd Test. More on this was discussed in Chapter 3, Section 3.5.

In the 2nd Test, participants were not asked to choose whether a scenario was Phishing or not because all the 7 scenarios were Phishing scenarios. Instead, they were asked to identify the behavioural factors involved in each of the Phishing scenarios by selecting the correct answer out of 7 possible predefined answers. The same 7 predefined answers were provided under the 7 scenarios with only one answer being the correct answer. Thus, participants could select one answer only. Each answer comprised two or three behavioural factors. Figure 4 shows one of the 7 scenarios on the 2nd Test with the 7 predefined answers listed below.



The following 7 scenarios are Phishing scenarios. You are requested to read them carefully and select from the list below each scenario the most suitable answer to describe the possible combination of human behavioral incentives used by the attacker to lure the victim in the scenario.

*** 3. Scenario 1: You have received an e-mail from the University Administration where you are registered on a full tuition basis as an online student, informing you about a raffle draw they are conducting to award winning students with huge discounts on payment that might reach 99%. You are invited to click a link with your student ID to confirm your identity and register for the draw. They are also telling you that the registration in the raffle draw is limited to the first 10 comers only, so take your chance now and do not wait. The prizes are a big surprise that you will find out if you click the link provided within the e-mail body.**

This is a Phishing situation wherein at least a combination of two human behavioral incentives were used by the phisher to lure the victim. Please select the correct answer from the list below:

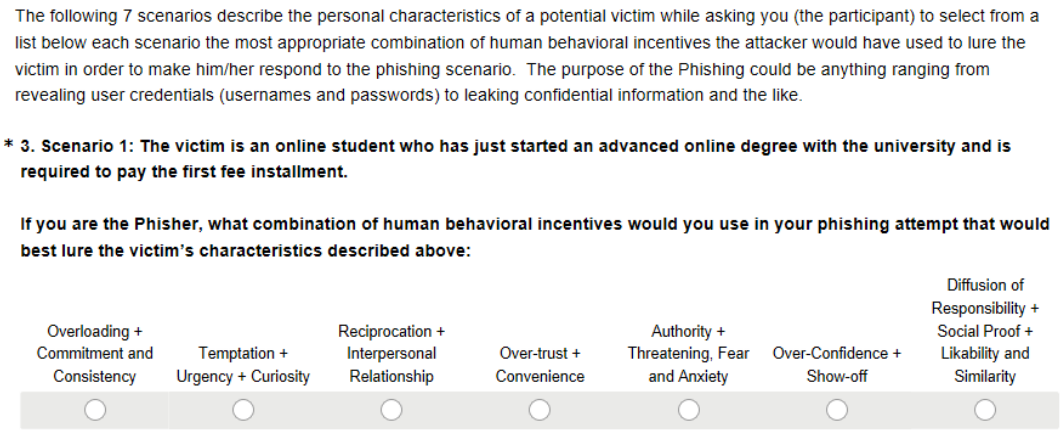
Overloading + Commitment and Consistency	Temptation + Urgency + Curiosity	Reciprocation + Interpersonal Relationship	Over-trust + Convenience	Authority + Threatening, Fear and Anxiety	Over-Confidence + Show-off	Diffusion of Responsibility + Social Proof + Likability and Similarity
○	○	○	○	○	○	○

Figure 4 Example of 2nd Test's Scenarios

SurveyMonkey Survey - the 3rd Test: After the same groups of participants had taken the 2nd Test, their scores on the 2nd Test represented the control group's scores against which their scores on the 3rd Test were compared (Experimental Group). So after the second awareness session had been attended, the 2nd Test became the pre-awareness assessment test and the 3rd Test became the post-awareness assessment test. The 3rd Test aimed to assess whether there was another significant improvement in participants' understanding about Phishing after attending the second awareness session compared to the assessment results from the 2nd Test.

Like the 2nd Test, the 3rd Test also comprised 7 Phishing scenarios combining two or three behavioural factors from the 16 factors to bait the victim. Hence, the 3rd Test also covered all the 16 behavioural factors using 7 scenarios only. The 3rd Test adds an extra level of difficulty to the scenarios while maintaining similarity of content with the 1st and 2nd tests. The extra level of difficulty in the 3rd Test was meant for testing the highest level of understanding acquired by the participants after attending the first and second awareness sessions. According to the adopted Situation Awareness Model (Endsley, 2015), the third level of understanding assessed by The 3rd Test is the 'Projection' level. In order to quantitatively measure the participant awareness improvements gained between the 2nd Test and the 3rd Test after attending the two awareness sessions, the participants' scores on each test were factored by the given weights of difficulty respective to the level of awareness level assessed by each test. More on this was discussed in Chapter 3, Section 3.5.

In the 3rd Test, participants were presented with 7 Phishing scenarios and were asked to infer the most relevant behavioural factors that a Phishing attacker would exploit according to the situation presented in each Phishing scenario and answer by selecting the correct answer out of 7 possible predefined answers. The same 7 predefined answers were provided under the 7 scenarios with one answer only being the correct answer where each answer comprised two or three behavioural factors. Figure 5 shows one of the 7 scenarios on the 3rd Test with the 7 answers below.



The following 7 scenarios describe the personal characteristics of a potential victim while asking you (the participant) to select from a list below each scenario the most appropriate combination of human behavioral incentives the attacker would have used to lure the victim in order to make him/her respond to the phishing scenario. The purpose of the Phishing could be anything ranging from revealing user credentials (usernames and passwords) to leaking confidential information and the like.

*** 3. Scenario 1: The victim is an online student who has just started an advanced online degree with the university and is required to pay the first fee installment.**

If you are the Phisher, what combination of human behavioral incentives would you use in your phishing attempt that would best lure the victim's characteristics described above:

Overloading + Commitment and Consistency	Temptation + Urgency + Curiosity	Reciprocation + Interpersonal Relationship	Over-trust + Convenience	Authority + Threatening, Fear and Anxiety	Over-Confidence + Show-off	Diffusion of Responsibility + Social Proof + Likability and Similarity
○	○	○	○	○	○	○

Figure 5 - Example of 3rd Test's Scenarios

The answers collected from the three tests were nominal data which was unsuitable for conducting quantitative analysis (Cohen, Manion and Morrison, 2011). Therefore, the collected nominal data was converted to parametric ratio data by calculating the sum of the correct answers for each question. This technique provided the required data conversion from nominal to ratio rendering the data suitable for consumption by quantitative parametric data analysis tests in order to statistically compare the variances between the mean scores among the three tests. The conducted variance analysis for the three tests' scores helped verify the learning outcomes and validate the effectiveness of the proposed conceptual framework used in conducting the awareness sessions as it also provided answers to Research Question 2. More on this will be discussed in detail in Chapter 5. Next, the two awareness sessions as tools for conducing awareness are discussed.

4.10 Awareness Session #1 and #2

Participants attended and interactively participated in the group discussions of the 1st and 2nd awareness sessions after they had completed the 1st and 2nd assessment tests respectively. That was to assess whether participants had progressed from the Perception level to the Comprehension level after the 1st awareness session and from the Comprehension level to the Projection level after the 2nd session. During these 2 awareness sessions about the behavioural factors underlying Phishing attacks, participants discussed in depth the scenarios from the 1st and 2nd Tests as examples to enhance their understanding of how the 16 behavioural factors would be used by attackers. The group discussions were important for participants to improve their awareness and take their awareness to the next level and gain the required knowledge for the post-awareness assessment tests.

Both awareness sessions were conducted online repetitively using Google hangouts to accommodate groups ranging from 5 to 10 participants until all the 40 participants categorized as the Online Group were covered. On the other hand, Face-to-Face version of the two sessions were also conducted using a classroom-based setup to accommodate the remaining 60 participants categorized as the Face-to-Face group. The agenda of the 2 awareness sessions consisted of a thorough review of the 16 behavioural factors followed by a group discussion of the imaginary Phishing scenarios on the 1st and 2nd assessment tests; making links to the 16 behavioural factors. The objective of the assessment tests was to verify that the learning outcomes participants demonstrated during the two awareness sessions were significantly reflected in participants' scores.

4.11 Data Collection Process

Data collection and participants' recruitments were carried out in parallel spanning over 8 months of coordination, arrangements, scheduling, re-arrangements and rescheduling. As participants were recruited, they were informed that the 5 step process had to be fully completed in order for their participation to be accepted and to successfully finish their participation cycle. This means that if they agreed to the participation consent, they were required to complete the 5 steps of participation already explained in Section 4.7.

When a participant had completed the 1st Test, an average of one week elapsed before he or she could join a group of around 5 to 10 participants to attend the online Awareness Session #1. When the awareness session was scheduled, an invitation was sent to each participant in the session attendee list for acceptance. Once invitations were accepted, the session was conducted online as scheduled using Google Hangouts. As more participants completed the 1st Test, they were joined and scheduled to attend Awareness Session #1.

Once a group of participants attended Awareness Session #1, they were sent a link to the 2nd Assessment Test and requested to complete it. When a group of participants had completed the 2nd Test, around a week later they attended Awareness Session #2. Once, a group had attended Awareness Session #2, attendees were sent a link to the 3rd Assessment Test. Finally, when a group of participants complete the 3rd assessment Test, their participation come to an end. This cycle continued almost for 20 weeks until all 40 online participants had completed the 5 steps of their participation. Although the process seems straight forward, it involved a lot of coordination and rescheduling for make-up sessions as some participants did not show up in the online awareness sessions, did not complete tests on time or rejected invitations due to time-zone issues or other personal reasons.

With the 60 Face-to-Face participants, the same cycle of steps 1, 3 and 5 took place. However the two Face-to-Face awareness sessions were conducted one time each for all 60 participants in a classroom based setup.

Both the Online and Classroom-based awareness sessions took about 1 hour and a half. However, the classroom-based sessions were more interactive compared to the online versions. In the classroom-based Face-to-Face awareness sessions, participants were asked to come out in pairs in front of all the other 60 participants and play roles of the Phishing victims and attackers in the imaginary phishing scenarios discussed during the session. This level of interactivity could not be easily achieved in the online sessions. Moreover, participants showed their creativity by coming up with their own phishing scenarios that were discussed and analysed as well during the group discussions in both the online and Face-to-Face sessions. Participant recruitment formed a critical part in how the data collection work progressed. Later in this chapter, participant recruitment is discussed.

4.12 Data Collection Issues

The data collection activities having been tightly coupled with recruitment of participants and participation became an over-whelming and time consuming process that demanded a lot of coordination effort and arrangements with participants. Eight months of hard work was put into finding the right candidates to participate in the project and to maintain their continued participation until all the 5 steps of the data collection process were successfully completed, a number of obstacles were encountered and eventually overcome along the way. The following are these obstacles dealt with during the course of data collection:

- Discontinued Participation: The participation process is a lengthy 5 step process that required the completion of all 5 steps. On average each participant's activities were planned and scheduled over two weeks however most participants would not be able to continue all the way and would just drop out part way through. As a result, more than 650 participants were dropped from data analysis. This necessitated a further search for more new participants every time it was thought that the required number of participants had been finally obtained. However, data gathering ceased when 100 participants successfully completed all the required steps.

- Diverse Time Zones Issues: Participants were located in different and diverse time zones as they were scattered in different parts of the world. Some participants were in Saudi Arabia, while others were in the US and the UK. This diversity of locations made it difficult to reconcile one standard schedule for conducting the online awareness sessions without some participants rejecting the invitation due to the inconvenient timing according to their time zones or simply not showing up at the scheduled session time after they had accepted the invitation. This situation necessitated rescheduling of the online awareness sessions over and over again. That, in addition to conducting make-up sessions at a very late Saudi Arabia time such as 2:00 a.m. to make up for absentees in order to help them synchronize their participation with the groups that had already been scheduled and conducted. The compulsory chronological order of the participation process also made it even more difficult to arrange and manage those make-up sessions since for example, participants were not supposed to attend Awareness Session #2 if they had not attended Awareness Session #1 and could not attend Awareness Session #1, had they not completed the 1st Test and so on. However, with more participants on board and more perseverance, this issue was finally overcome.
- Participants' Slack Commitment to Schedule and Deadlines: 100 of 750 participants had completed their participation cycle to the end. The 100 participants who made it to the end, had also issues with their commitments to the set deadlines. For example, not showing up for the online awareness session, which happened very often, resulted in having to repeat parts of the session every time a late comer joined the discussion in order to get them up to speed. The learning outcomes of late comers may have also been affected by their tardy attendance (Gillies and Quijada, 2008) which disturbed the flow of the session and might have potentially affected their scores on the three tests and the overall objectives of those awareness sessions.

- Technical Limitations of Online Environments: Last but not least, the limitations of online environments for conducting group discussions may have had implications for how effectively the group discussions were conducted and the objectives achieved (Johnson et al, 2000; Wang and Woo, 2007; Walsh and Brown, 2013). One of the major restrictions of online environments encountered was the limited number of participants to accommodate in an online session. For instance, Google Hangouts has a limitation of 10 participants per session (Isaacson, 2013), while Skype has a limit of 15 participants (Nyíri, 2008) per session. This made it impossible for the group discussions to accommodate more participants in fewer sessions to save time and coordination effort, and lessen the need for conducting more make-up sessions. Another limitation was the diverse online supporting infrastructure from one location to another e.g. online technologies used such as the graphical media based collaboration web-based applications which are band-width dependent (Zhang, Zhou, Briggs, and Nunamaker, 2006) and hence are heavily affected by poor bandwidth connections. Poor bandwidth forced some participants to disconnect from the session and reconnect multiple times thereby wasting session time and interrupting the flow of the discussion. These disconnected participants were sometimes thrown out of the session without being able to reconnect resulting in having to reschedule them for a make-up session or two in the future thus requiring further coordination, arrangements, more time and effort. Another limitation was the participants' detachment from the online group discussion in case the moderator failed to maintain adequate social presence (Gunawardena, 1995). This detachment happens because the participants may be too conscious of the actual physical barriers and the distance dividing them from the group they are engaged with over a virtual connection. Unlike the online group discussion, the Face-to-Face group discussions participants experienced physical presence, eye contact and facial expressions, which highly contributed to strengthening their feeling of social presence and therefore engagement in the discussion (Gunawardena, 1995).

In the next sections, the tests' scores standardization and tests' difficulty levels weighting are discussed.

4.13 Standardizing Scores

As the 1st Test has 20 questions, its score scale is 20 compared to the 2nd and 3rd tests which both have only 7 questions and a corresponding score scale of 7. In order to statistically compare the scores between the 3 tests and measure the significance of the differences in participant awareness, the score scales of the 3 tests had to be reconciled and therefore standardized to the scale of 7. In other words, the scores of the 1st Test were scaled down from 20 to 7 in order to match the scoring scale of the 2nd and 3rd tests. The 1st Test's scores were standardized according to the following formula (IBM, 2016; Aiken, 1987) taking the following steps:

Standardizing Score Scale: Standardization of the 1st Test Scores was achieved by recalculating the original score as in the following formula: (7 - 1) * ((The 1st Test Original Score) -1) / (20 - 1) + 1. This is interpreted as the following:

1. Upper bound of the 1-7 scale = 7 is subtracted from the (Lower bound of the 1-7 scale = 1) which is equal to (7-1 = 6).
2. Then, the Original Score is subtracted from the (Lower bound of the 1-7 scale = 1).
3. Then the results of the two subtractions above are multiplied together.
4. Then, the product of the multiplication above is divided by the result of subtracting the (Upper bound of the 1st Test Original Score's Scale 1-20:= 20) from the (Lower bound of the 1st Test Original Score's Scale 1-20:= 1) which is equal to (20-1 = 19).
5. Then, adding the result of the division above to the (Lower bound of the 1-7 scale: = 1).

For example, the original score of 13/20 is scaled down to 4.79/7 by recalculating it as (7-1)*(13-1)/ (20-1) +1 = 4.79 following the steps below:

- (7-1) = 6
- (13-1) = 12
- 6*12 = 72
- 72 / (20-1) = 3.789
- 3.789 + 1 = 4.789

The standardized scores on the 1st Test were used in all statistical tests that involved comparison of any two different sets of participants e.g. the Online and Face-to-Face participant groups within the same test. However, for statistical tests that required comparison between the scores of two different tests whose level of difficulty varied, the standardized scores needed also to be weighted to equate the differences between the incremental levels of difficulty between the tests.

Finally, not all the data collected for analysis are represented as test scores which are ratio data type (Cohen, Manion and Morrison, 2011). Some of the collected data are Likert-scale items such as the participants' age groups, education levels, which are ordinal data type (Cohen, Manion and Morrison, 2011). Therefore, these ordinal data are converted to ratio data type using an appropriate coding scale based on their level (Allen and Seaman, 2007). For example, education levels like High School, Bachelor, Masters and PhD are coded as 1, 2, 3 and 4 respectively with High School given the lowest value of 1 and PhD given the highest value of 4 on a scale of 1 to 4. After the conversion, all data becomes ratio data and therefore are ready for statistical analysis. The next section discusses how the weighted scores are calculated.

4.14 Tests' Difficulty Levels

The Situation Awareness Model (Endsely, 2015) was adopted to raise awareness about the 16 Phishing behavioural factors to the participants. Thus, the 3 tests aimed to measure the participant awareness and understanding levels of how the 16 behavioural factors would be exploited in 16 imaginary Phishing scenarios. It was assumed that participant awareness would improve gradually from the initial Perception level, to the Comprehension level and finally to the Projection level as a result of participants attending the awareness sessions between the tests. Therefore, the role of each test was to measure each level of the Endsley's Situation Awareness (Endsley, 2015) by quantitatively analysing participant awareness improvements reflected by the participants' scores on each test. Despite the content of the 3 tests being similar, the questions and the answering approach were setup slightly different. Each test was designed with increasing difficulty matching the corresponding difficulty level incorporated in the questions.

To maintain objectivity and fairness in the scoring scheme for the 3 tests, adjustment weights (Morris, 1982; Braun and Holland, 1982) were added to equate the differences in the scores due to the gradual difficulty levels on each test. Therefore, in order to assess the significance of participant awareness improvements in each test, a proper weight was assigned to the difficulty level on each test and multiplied by each participant’s score. The results then are the adjusted and reconciled scores between the three tests. Next, the score weighting mechanisms are discussed.

4.15 Weighting Scores

Tests scoring scheme must be fair and should consider the variability in the difficulty levels between the three tests when scores are compared regardless of the differences in test forms (Kolen and Brennan, 2004). Therefore, assigning a proper weight to factor in the level of difficulty in the scoring scheme is suggested (Kalliniatis et al, 2017; Poepjes, 2012) to ensure fair comparison between test scores.

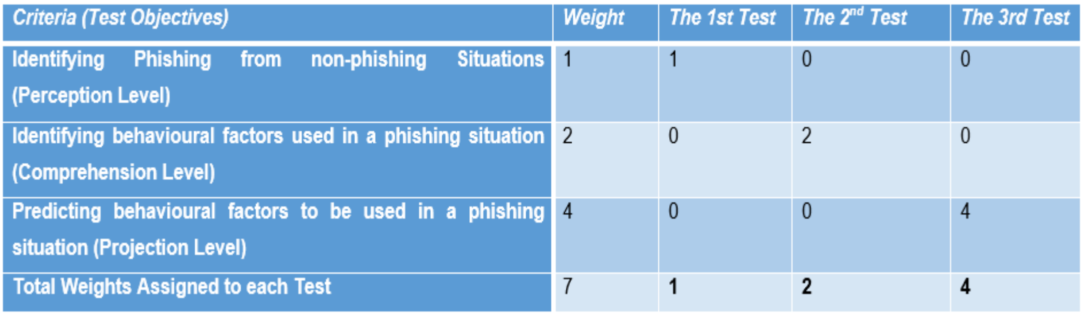
Following in the steps of the Situation Awareness Weighted Model (SAWN) model (Kalloniatis, Ali, Neville, La, Macleod, Zuparic, and Kohn. 178-196) to ensure that the weights assigned to each level of difficulty were objective, a consistent list of criteria (Jadhav and Sonar, 2009) was developed based on the answering approach required by each test. Each criterion is assigned a weight based on the awareness level (Endsley, 2015) measured. Then, each test was compared to every criterion and assigned a value which then was added up to meet the preset criteria for that level. If the test did not meet a criterion, a value of 0 was entered. The summation of all criteria values for a test was the final weight assigned to the test’s difficulty level. Table 1 shows the criteria list and the calculated weight for each test.

Table 1 - Criteria for Calculating Weights for Tests

Criteria	Weight	The 1st Test	The 2ndTest	The 3rd Test
Identifying Phishing from non-phishing Situations	1	1	0	0
Identifying behavioural factors used in a phishing situation	2	0	2	0
Predicting behavioural factors to be used in a phishing situation	4	0	0	4
Total Weights Assigned to each Test	**7**	**1**	**2**	**4**

As such, determining the proper weight to assign to a test, used a weighted objectives table (WOT) as a means of comparing several different alternatives ranked based on a list of criteria (Caro, 2011; Pesonen et al, 2001). Also, Wright, Taekman and Endsley (2004) used objectives as a means to measure Situation Awareness. In the same way, WOT is used in this research to compare the different difficulty levels between the three tests and to rank them accordingly. The most difficult test is assigned the highest weight, while the least difficult test is assigned the lowest weight as illustrated in Table 2.

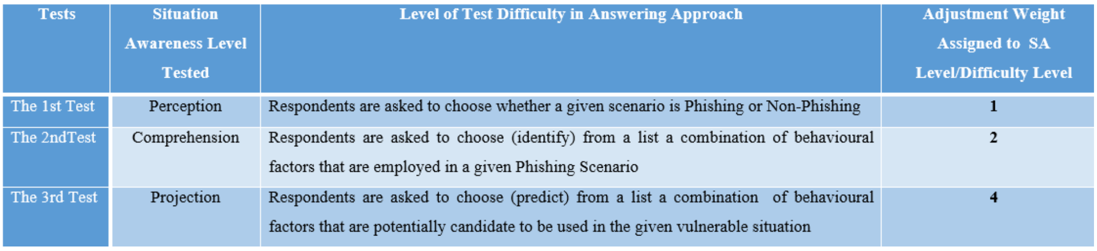
Table 2- Weighted Objectives Table for Tests



Criteria (Test Objectives)	*Weight*	*The 1st Test*	*The 2nd Test*	*The 3rd Test*
Identifying Phishing from non-phishing Situations (Perception Level)	1	1	0	0
Identifying behavioural factors used in a phishing situation (Comprehension Level)	2	0	2	0
Predicting behavioural factors to be used in a phishing situation (Projection Level)	4	0	0	4
Total Weights Assigned to each Test	7	**1**	**2**	**4**

Thus, the weights 1, 2 and 4 were proportionally assigned to the 3 Situation Awareness Levels namely Perception, Comprehension and Projection respectively as illustrated in Table 3.

Table 3 - Weights Assigned to Situation Awarness 3 Levels



Tests	**Situation Awareness Level Tested**	**Level of Test Difficulty in Answering Approach**	**Adjustment Weight Assigned to SA Level/Difficulty Level**
The 1st Test	Perception	Respondents are asked to choose whether a given scenario is Phishing or Non-Phishing	**1**
The 2ndTest	Comprehension	Respondents are asked to choose (identify) from a list a combination of behavioural factors that are employed in a given Phishing Scenario	**2**
The 3rd Test	Projection	Respondents are asked to choose (predict) from a list a combination of behavioural factors that are potentially candidate to be used in the given vulnerable situation	**4**

Considering the gradually increasing difficulty levels in each test's objectives in proportion to the hierarchically ascending SA levels, respective ascending weights are assigned to each test based on the result of comparing and ranking the three Tests' objectives. These Tests' objectives provide the basic criteria for such comparison. For example, the 1st Test's objective is to assess participants' (Perception Level) ability by simply asking them to recognize a Phishing scenario without assessing their knowledge of the underlying behavioural factors exploited in the Phishing scenario. Hence, the Perception level was given the weight value of 1; compared to the 2nd Test's objective which aims to assess the 'Comprehension' level by requiring participants to take one step further to recognize the behavioural factors exploited in the Phishing scenario. Hence, the Comprehension Level as given the multiple value of 1 which is 2. Finally, the 3rd Test which aims to assess the 'Projection' level by asking participants to demonstrate their ability to predict the behavioural factors most exploitable in the Phishing scenario was given the multiple value of 2 which is 4. Thus, the multiple weights are calculated and assigned to the tests according to the different levels of difficulty embedded in each Test's response objective. Hence 1, 2 and 4 are used as the weight factors to effectively account for the respective tests' levels of difficulty when comparing scores between these tests.

There is no absolute way to assign specific weights to precisely match up to the difficulty level (Wang and Stanley, 1970) on each test since the judgement of how difficult a test is relative and judgemental. Any combination of weights on different scales that represent ascending hierarchical values could be used and actually have been experimented with. Nevertheless, the weight values (1, 2 and 4) are chosen based on its suitability to the standardized scoring scale of 1-7. Accordingly, the three tests' maximum standardized score will be 7 out of 7 and the total weights assigned to the three tests are also 7 as shown in Table 3.

As a result of this setup, a weight of one (1) is assigned to the 1st Test which provides assessment for the SA level 1 ‘Perception’. The weight of one will ensure that the same score is rendered after multiplying the score by the given weight e.g. one (1). This is to ensure that the starting difficulty level will match the ‘Perception’ level. In other words, the weight assigned to each test will be based on multiples on a scale of 1-7. As a result, The 1st Test will be assigned the weight 1 out of 7, which will always return the actual score e.g. multiplying the actual score by 1, since it represents the preliminary awareness Level – “Perception”. Then, 6 units on the weighting scale are left after the ‘1’ was assigned to the 1st Test. Then, the 2nd Test is assigned double the weight given to the 1st Test which will be 2 out of 7. In the same way, the 3rd Test will be assigned double the weight given to the 2nd Test, which is 4. Adding the weights assigned to the three tests together (1+2+4) results in a total of 7 out of 7. If the scores of the three tests are added together to be equal to 100% and then the total is divided by 7, the 1st Test will take 1 unit out 7, the 2nd Test will take 2 units out 7 and the 3rd Test will take the remaining 4 units on the weighting scale.

The weights (1, 2 and 4) are converted to percentiles of the sum of the three Tests’ full weighted scores which is equivalent to the weighting scale of 7, This conversion is accomplished by dividing the full percentage (100%) by the total weighted score which is (7), resulting in the value of 14.29 for each unit on the weighting scale of 7. To calculate the percentile equivalent to each test’s assigned weight, the respective test’s assigned weight value is multiplied by 14.29. For example, the weight of ‘1’ will be calculated as (1* 14.29 = 14.29), the weight of 2 will be calculated as (2 * 14.29 = 28.57), and finally the weight of 4 will be calculated as (4 * 14.29 = 57.14). So, the final weighed score will be the result of multiplying the actual score by the respective test’s percentile and dividing the product of the multiplication by 100. This means if a participant scores a 7 out of 7 on the 1st Test, his or her weighted score will be (7 * 14.29) / (100), whereas the same score of 7 on the 2nd Test will be calculated as (7 * 28.57) / (100) and of course the same score on the 3rd Test will be calculated as (7 * 57.14) / (100).

In other words, the total weighted scores a participant has accumulated over all the three tests will be calculated as one final score weighted over the scale of 7 which is equal to the sum of full scores on the three tests combined. For example assuming a participant had scored a full mark of 7 on all the three tests, his weighted full score on the 1st Test' would be 1 out of 7, his weighted full score on the 2nd Test would be 2 out of 7, while his weighted full score on the 3rd Test would be 4 out of 7. Adding these three weighted full scores (1+2+4) will be the participant's final weighted score of 7 out of 7 over the range of the three tests combined. Hence, if a participant has the same score on the 3 tests, each score will be evaluated according to the level of difficulty manifested by each test and the weight assigned to each test. This way, the scoring scheme is fairer considering the varying difficulty levels on each test. Table 4 shows these weights and their equivalent percentiles and how they are used to calculate the final weighted scores assuming the actual score is a full mark of 7 out of 7 on each test.

Table 4 - Distribution of Weights

Tests	Actual Scores	Assigned weights on the scale of 7	Weighs % of the scale	Weighted Scores
The 1st Test	**7**	1	14.29	**1.00**
The 2ndTest	**7**	2	28.57	**2.00**
The 3rd Test	**7**	4	57.14	**4.00**
Final Score	**21**	**7**	**100**	**7.00**

There is also another benefit from the adopted weighting scale of 7 as shown in Table 4. Using the weighting scale, the ISACM (Poepjes, 2012) the Information Security Awareness and Capability Model can be easily adopted and implemented. The ISACM consists of three components:

1. Risk Importance is assigned in this study the maximum value on the weighting scale which is 7. This means that the objective of our adopted SA conceptual framework (SAMFP) is to raise participant awareness to reach the maximum of the weighting scale by scoring a full mark on all the three tests.
2. Risk Capability which is the current level of awareness a participant achieves after attending the awareness sessions. This is represented by the participants' scores on each test as shown in Table 4

3. Risk Residual, which is the difference between the Risk Importance (assigned the value of 7) minus the participant's current awareness capability represented by the participant's final weighed score after taking the three tests.

The Risk Level is calculated based on the risk residual (difference between Risk Capability and Risk Importance). For example, if the Risk Residual is less than or equal to the sum of the weights assigned to the assessed SA Level and the levels below it, the participant's Risk Level will be categorized under that respective SA level represented by the test. In other words, if the Risk Residual is equal to or less than 1 (Weight assigned to 1st Test), the Risk Level will be categorized as 'Perception Level Risk'. Similarly if the Risk Residual is greater than 1 and equal to or less than 3 (Sum of (1+2), the weights assigned to the 1st and 2nd Tests respectively), the Risk Level will be categorized as "Comprehension Level Risk" and finally, if the Risk Residual is greater than 3 up to 7 (the sum of (3+4), the weights assigned to the 1st, 2nd and 3rd Tests respectively) the Risk Level will be categorized as "Projection Level Risk".

These Risk Level categorizations are meant to distinguish participants from an awareness level risk standpoint. For example, a participant whose category is in the 'Comprehension Level Risk' means that he/she has successfully acquired a Risk Capability beyond the 'Perception Level', but yet has to acquire more awareness and knowledge to successfully achieve more awareness improvements within the range of the same level capability e.g. the "Comprehension Level". While a participant whose category is in the 'Projection Level', has successfully acquired a Risk Capability of awareness beyond the Perception and Comprehension levels as that would adequately protect him or her from the risks within the achieved awareness levels, yet they still need more awareness sessions to reach the maximum capability within the range of the their existing level e.g. the Projection level. Thus, the participant is given more awareness sessions to help him/her achieve the advanced "Projection" SA level capability. Even if a participant achieves the highest Risk Capability which is 7 on the adopted weighted score scale, he or she would still be within the 'Projection Level Risk'. There is always a potential chance for risk which means that risk cannot be fully eliminated, but rather be reduced through continued awareness efforts.

Adopting the ISACM (Poepjes, 2012) as part of this research’s conceptual framework (SAMFP) has enabled not only measuring the current level of Situational Awareness participants have achieved, but also quantifying the progress made by participants over time and the gap left to be filled by performing more cycles of awareness sessions and tests to help participants achieve the maximum objective and Risk Importance level targeted.

Finally, according to the above weights assigned to each SA awareness level and their respective tests, the tests’ scores are calculated to ensure the comparison between scores are fair and that the varying levels of difficulty on each test is factored fairly in the calculation of the final resulting score. In Chapter 5, the scores from the 3 situation awareness level assessment tests are statistically compared and analysed for variance significance.

The weighed scores are only used when comparing scores between two tests or more in which the level of difficulty varies. However, when comparing scores between two sets of participants e.g. the Online and Face-to-Face participant groups within the same test, the actual unweighted scores will be used instead.

Chapter 5 Results

This chapter aims to present the results of the data analysis and highlight the findings from the hypotheses testing and the correlations identified between the different datasets and participants' demographics. This chapter presents the results of the research into the effectiveness of the conceptual model SAMFP in improving participant awareness of the behavioural factors used in Phishing attacks. In this chapter, analysis of the scores of the 3 assessment tests will be discussed covering the distribution of scores by question where each question is linked to the identified 16 human behavioural factors. The questions which received the most number of correct answers as opposed to those with the least number of correct answers will also be analysed and discussed. After that, the discussion will cover the 18 hypotheses and the testing procedure. In appendix A, the inferential analysis procedure developed for analysing the 18 hypotheses is discussed in detail. Furthermore, the correlations of the hypotheses' analysis results with the results from the descriptive analysis of participants' demographics and tests' questions will also be elaborated on. Finally, the discussion will conclude with a detailed analysis of the characteristics of the highest and lowest ranked participants in the 3 assessment tests.

5.1 Analysis of 1st Test's Scores – Preliminary Awareness (Perception)

The 1st Assessment Test features 16 Phishing scenarios out of 20 where each scenario potentially employs one of the 16 behavioural factors as the main incentive or bait set for trapping a victim. Participants have to answer whether the scenario is a Phishing or non-Phishing scenario based on their preliminary understanding of the scenario. If they answer the scenario as Phishing, they are then asked to describe in their own words the reasons for choosing such an answer. These reasons are compared to the actual behavioural factor underlying the scenario. This is to see how close and relevant their answers are to the actual behavioural factor in the scenario and to evaluate which of the 16 behavioural factors are more easily recognized by the participants. Choosing a correct answer on the 1st Test does not necessarily mean the participant recognized the correct behavioural factor underlying the scenario. The participant could have just followed his/her intuitive feelings about something phishy in the scenario without actually knowing the actual behavioural factor exploited in the scenario. This is left for the two awareness sessions to educate and equip participants with in-depth knowledge about these behavioural factors.

Only 100 of the 750 participants continued their participation. Consequently, the scores of the 100 participants who completed the full cycle of participation are analysed in order to measure awareness improvements and quantify risk residuals between attended awareness sessions and tests through comparing the 100 participants' three tests' scores.

5.3 1st Test's Scores for 100 Participants

The scores of the 100 participants who completed the full cycle of participation are now analysed. The 4 non-Phishing scenarios had the highest number of correct answers as illustrated in Figure 8. Phishing Scenarios #3, 18 and 8 representing 'Over-Confidence and Self-Consciousness', 'Convenience' and 'Social Proof' were correctly answered by the highest numbers of participants 79/100, 74/100 and 71/100 respectively; followed by scenarios #16 and 1 representing 'Diffusion of Responsibility' and 'Temptation' which both were correctly answered by 66/100 participants. The other scenarios representing the remaining behavioural factors ranged between 63/100 and 28/100 participants. Scenarios #11 and 4 representing 'Reciprocation' and 'Over Trust' respectively are the last two scenarios which received the lowest number of correct answers. See Figure 6 for further details.

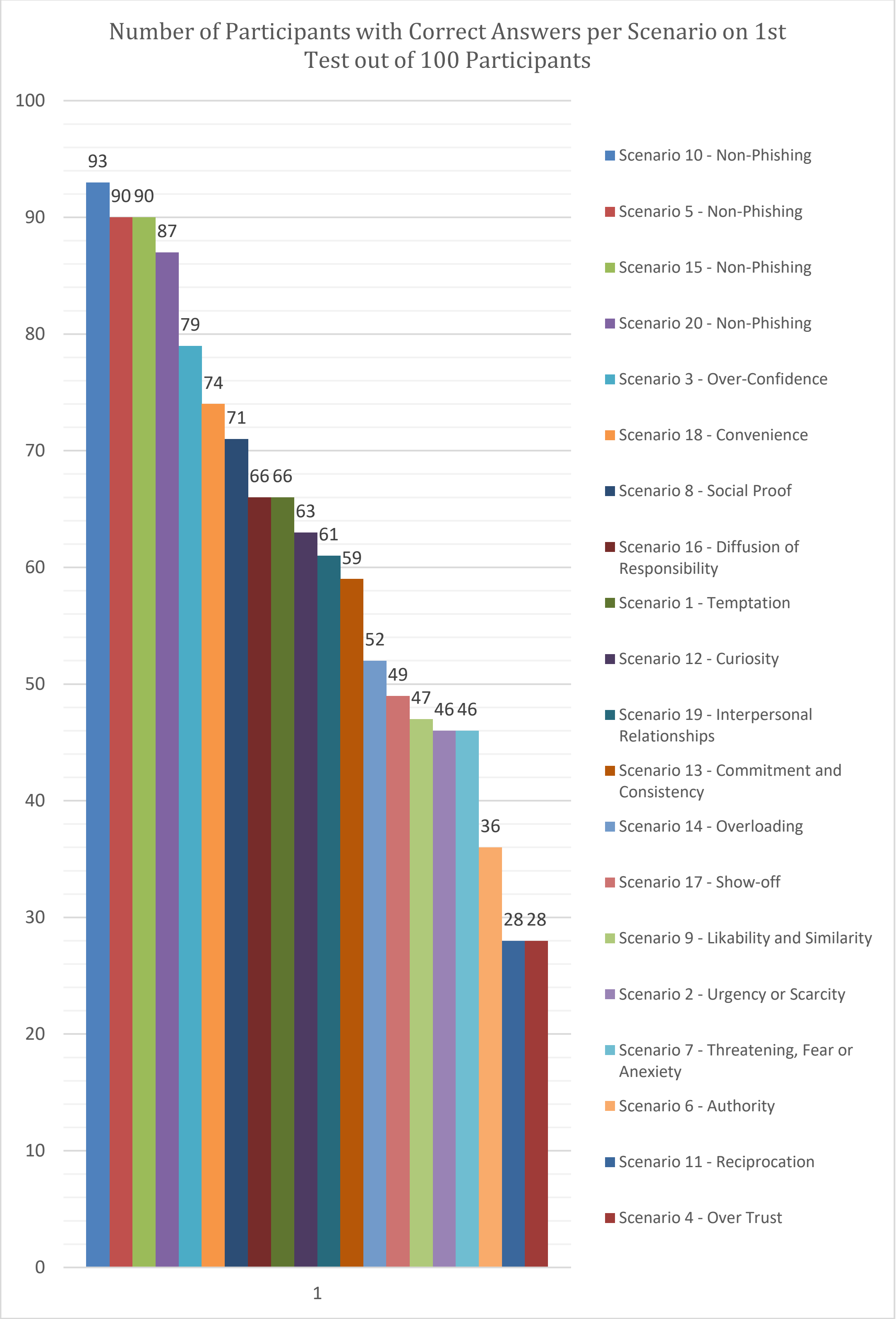
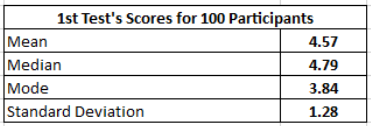


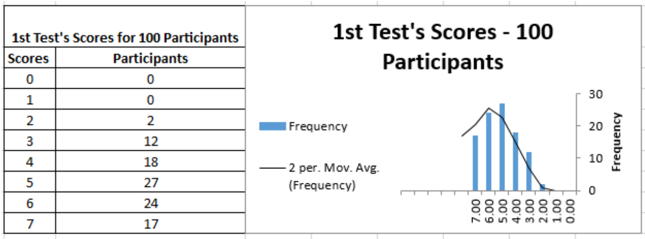
Figure 6-Number of Participants with Correct Answers per Scenario on 1st Test out of 100 Participants

From a score standpoint, the average score recorded for the 100 participants on the 1st Test was 4.57 as shown in Table 9. Figure 6 shows the distribution of the 100 participants over the 1st Test's score scale. Thus, most of the 100 participants' scores are distributed between 5 and 6 over 7, namely 27% participants scored 5/7, 24% scored 6/7 and 18% scored 4/7. On the other hand, only 17% participants scored 7/7.

Table 9 - Statistics of 1st Test's Scores – 100 Participants



1st Test's Scores for 100 Participants	
Mean	**4.57**
Median	**4.79**
Mode	**3.84**
Standard Deviation	**1.28**



1st Test's Scores for 100 Participants	
Scores	Participants
0	0
1	0
2	2
3	12
4	18
5	27
6	24
7	17

Figure 6 - 1st Test's scores for 100 participants

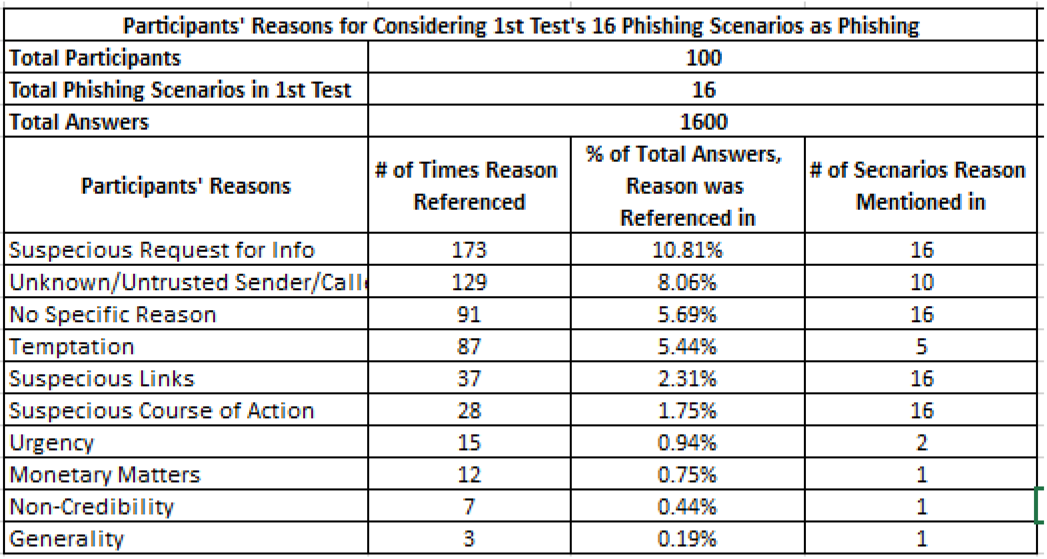
The previous statistics show that 100 participants share a similar trend of scores and level of awareness about Phishing in general and Phishing related behavioural factors in particular. This trend explains the preliminary level of awareness in the sampled participants where the average score is 4.31. Mapping the scores to the underlying behavioural factors in each scenario indicates that the 4 non-Phishing scenarios had the highest scores. On the other hand, 'Over-Confidence' and 'Convenience' were the most highly correctly answered scenarios out of the 16 Phishing scenarios, whereas, the 'Reciprocation' and 'Over Trust' scenarios were the least correctly answered Phishing scenarios. Next, the participants' reasons for considering a Phishing scenario as 'Phishing' are discussed.

5.4 1st Test's Participants' Comments (Reasons for Phishing)

The objective of analysing participants' comments is to evaluate the participants' preliminary awareness level (Perception) of the 16 behavioural factors prior to attending the training sessions. Participants' reasons for considering a scenario as Phishing are analysed to see how closely they came to describing the actual behavioural factors exploited in each Phishing scenario.

Table 5 and Figure 7 show the 100 participants' reasons are ordered by the highest frequency. Accordingly, 'Suspicious Request for Information' (173, 10.81%) and 'Unknown/Untrusted Sender/Caller' (129, 8.06%) are the most highly frequent reasons respectively. However, 'No Specific Reason' (91, 5.69%) was the third highest reason. On the other hand, 'Temptation' (87, 5.44%) was the fourth highest reason whereas 'Suspicious Links' (37), 'Suspicious Course of Action' (28), 'Urgency' (15), 'Monetary Matters' (12), 'Non-Credibility' (7) and 'Generality; (4) were the least frequently mentioned reasons respectively.

Table 5 – 100 Participants' reasons (entered comments) for ansering 1st Test's Scenarios to be Phishing, ordered by highest frequency



Participants' Reasons for Considering 1st Test's 16 Phishing Scenarios as Phishing			
Total Participants	100		
Total Phishing Scenarios in 1st Test	16		
Total Answers	1600		
Participants' Reasons	# of Times Reason Referenced	% of Total Answers, Reason was Referenced in	# of Secnarios Reason Mentioned in
Suspecious Request for Info	173	10.81%	16
Unknown/Untrusted Sender/Call	129	8.06%	10
No Specific Reason	91	5.69%	16
Temptation	87	5.44%	5
Suspecious Links	37	2.31%	16
Suspecious Course of Action	28	1.75%	16
Urgency	15	0.94%	2
Monetary Matters	12	0.75%	1
Non-Credibility	7	0.44%	1
Generality	3	0.19%	1

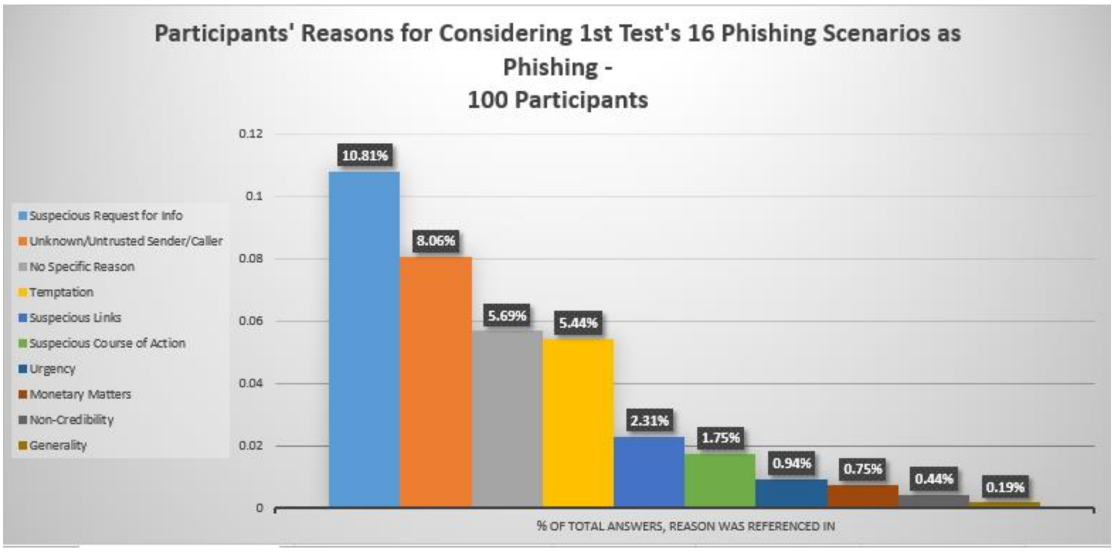


Figure 7 – 100 Participants' comments on 1st Test ordered by highest frequency of entries

Analysis of the participants' reasons reveal a similar level of awareness specifically about the 16 Phishing related behavioural factors. Their preliminary awareness level manifested the following observations from the comments' frequency analysis:

- Uncertainty denoted by the 'No Specific Reason' comment about the phishing reasons behind the 16 scenarios was common for both participant datasets and the most highly mentioned reason.
- Generality of expressing Phishing reasons which was denoted by the 'Suspicious Request for Info' and 'Suspicious Course of Action' comments mentioned in almost every scenario.

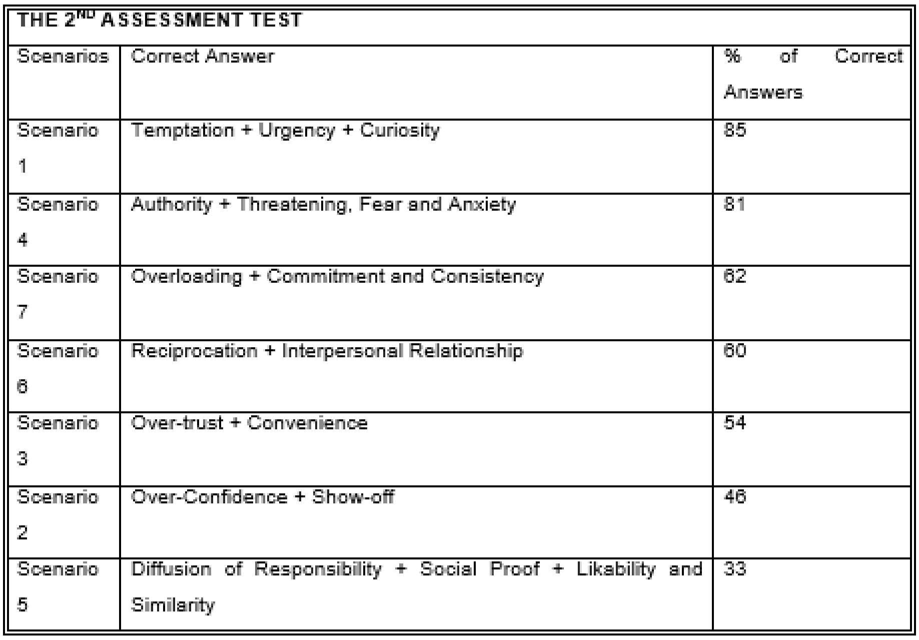
On the other hand, the only 2 out of the 16 behavioural factors that were correctly referenced in participants' comments were 'Temptation' and 'Urgency' which are represented by Scenarios 1 and 2 respectively. 'Temptation' was referenced by 38% of the participants ,while 'Urgency' was referenced by 6% only. The remaining 14 behavioural factors were not clearly and specifically referenced in participants' comments which indicates lack of in-depth awareness about these behavioural factors and their exploitation in Phishing scenarios.

5.5 Analysis of the 2nd and 3rd Tests' Scores (Comprehension and Projection)

Unlike the 1st Test, the 2nd and 3rd Tests are shorter featuring only 7 scenarios each; compared with the 20 scenarios on the 1st Test. However, these 7 scenarios are designed with more difficulty and complexity than those scenarios on the 1st Test where participants had to only choose an answer of whether a scenario was Phishing or Non-Phasing whereas in the 2nd and 3rd Tests' scenarios, participants have to identify the Phishing behavioural factor(s) underlying each phishing scenario. The reason for the increased difficulty is to evaluate and measure whether the participant awareness level has improved after attending the 1st awareness session and that participants have achieved the next levels of awareness namely Comprehension and Projection through the applied SAMFP model underpinning this study.

Each scenario on the 2nd Test introduces a potential Phishing situation where a combination of behavioural factors are employed to set up a potential victim. The participants should be able to identify the combination of these behavioural factors underlying each scenario in order for the expected improvement in their awareness level to be indicated by the test results. As illustrated by Table 6 and Figure 8, the highest score recorded was 85% for Scenario #1. This means that 85 participants out of 100 have correctly answered Scenario #1 choosing 'Temptation + Urgency + Curiosity' as the correct combination of behavioural factors employed in Scenario #1. On the other hand, Scenario #5 had the lowest score of 33% in which 'Diffusion of Responsibility + Social Proof + Likability and Similarity' was the combination of behavioural factors underlying this Scenario. Table 10 and Figure 10 show the 7 scenarios of the 2nd Test, their correct answers and the percentage of participants who correctly answered each question.

Table 6 - 2nd Test's Scenarios and % of Correct Answers



THE 2ND ASSESSMENT TEST		
Scenarios	Correct Answer	% of Correct Answers
Scenario 1	Temptation + Urgency + Curiosity	85
Scenario 4	Authority + Threatening, Fear and Anxiety	81
Scenario 7	Overloading + Commitment and Consistency	62
Scenario 6	Reciprocation + Interpersonal Relationship	60
Scenario 3	Over-trust + Convenience	54
Scenario 2	Over-Confidence + Show-off	46
Scenario 5	Diffusion of Responsibility + Social Proof + Likability and Similarity	33

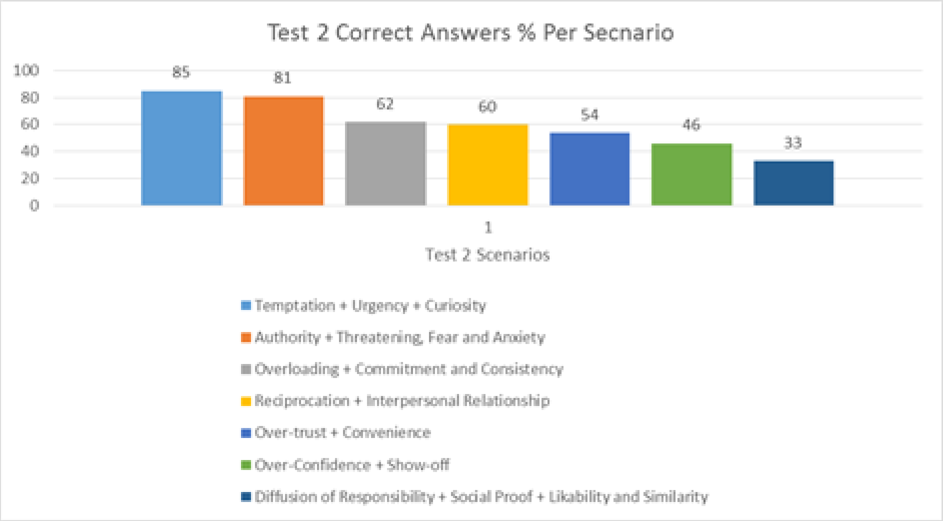
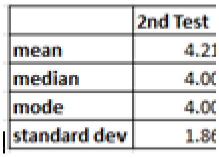


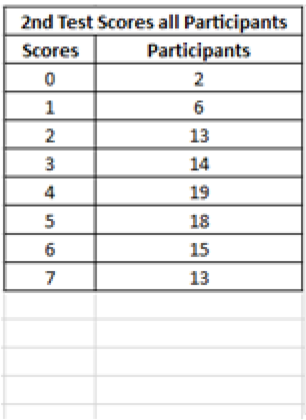
Figure 8 - 2nd Test's Scenarios and % of Correct Answers per scenario

Table 7 shows that the 2nd Test's unweighted scores have a mean score of 4.21 with a standard deviation of 1.86, Mode of 4.00, and a Median of 4.00. Distribution of the participants of the 2nd Test over the range of scores from 0 to 7 using the frequency bin is shown by Figure 9 where the highest number of participants is 19 scoring 4, followed by 18 participants scoring 5, 15 participants scoring 6, 14 participants scoring 3, 13 participants scoring 7, 13 participants scoring 2, 6 participants scoring 1 and finally only 2 participants scoring as low as 0.

Table 7 - 2nd Test's statistics



	2nd Test
mean	4.21
median	4.0
mode	4.0
standard dev	1.8



2nd Test Scores all Participants	
Scores	Participants
0	2
1	6
2	13
3	14
4	19
5	18
6	15
7	13

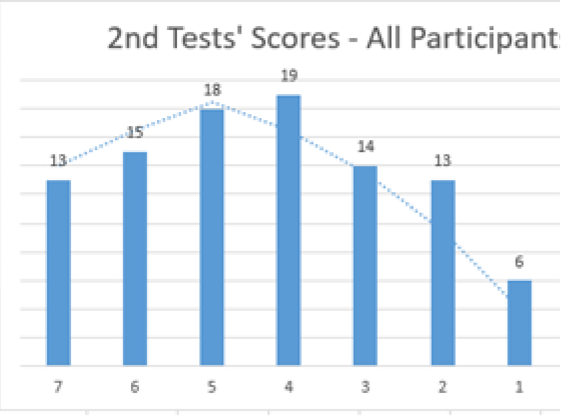
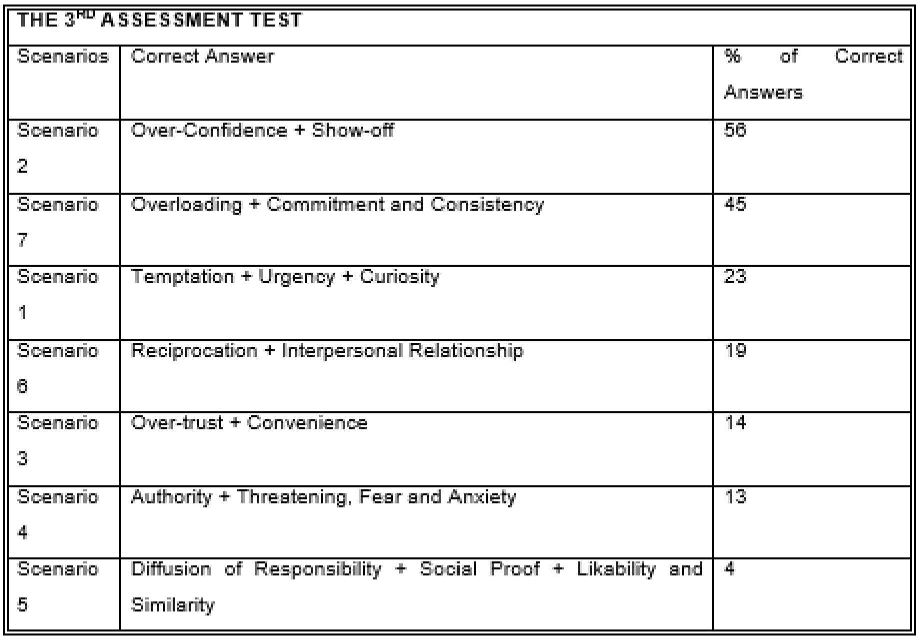


Figure 9 - 2nd Test's scores for all participants

Similar to the 2nd Test, the 3rd Test has also 7 scenarios designed with even more difficulty and complexity than those scenarios of the 2nd Test where participants have to identify the Phishing behavioural factor(s) projectable for exploitation in each phishing scenario based on projection (Endsley, 2015). The reason for the increased difficulty is to evaluate and measure whether the participant awareness levels have improved after attending the 2nd awareness session and that participants have achieved the ultimate level of awareness which is Projection by which participants can predict the potential behavioural factors a Phisher would exploit in a Phishing scenario.

As illustrated in Table 8 and Figure 10, the highest score recorded was 56% for Scenario #2 choosing ‘Over-Confidence + Show-Off’ as the combination of behavioural factors most exploitable by an attacker. On the other hand, Scenario #5 had the lowest score of 4% in which ‘Diffusion of Responsibility + Social Proof + Likability and Similarity’ was the combination of behavioural factors potentially projectable for Scenario #5. Table 9 and Figure 11 show the 7 scenarios on the 3rd Test and their percentage of correct answers.

Table 8 – 3rd Test's scenarios ordered by highest number of correct answers



THE 3RD ASSESSMENT TEST		
Scenarios	Correct Answer	% of Correct Answers
Scenario 2	Over-Confidence + Show-off	56
Scenario 7	Overloading + Commitment and Consistency	45
Scenario 1	Temptation + Urgency + Curiosity	23
Scenario 6	Reciprocation + Interpersonal Relationship	19
Scenario 3	Over-trust + Convenience	14
Scenario 4	Authority + Threatening, Fear and Anxiety	13
Scenario 5	Diffusion of Responsibility + Social Proof + Likability and Similarity	4

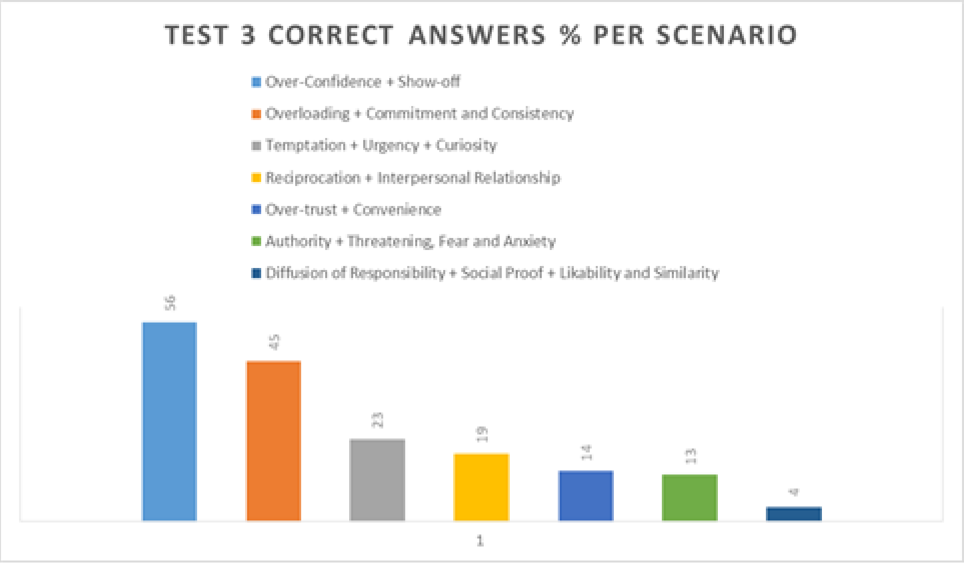
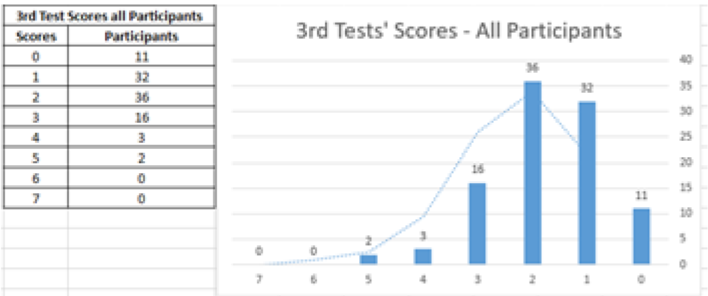


Figure 10 - 3rd Test's scenarios ordered by highest number of correct answers

Table 9 shows that the 3rd Test's unweighted scores have a mean score of 1.74 with a standard deviation of 1.08, Mode of 2.00, and a Median of 2.00. Distribution of the participants of the 3rd Test over the range of scores from 0 to 7 using the frequency bin is shown by Figure 11 where the highest number of participants was 36 participants scoring 2, followed by 32 participants scoring 1, 16 participants scoring 3, 11 participants scoring 0, 3 participants scoring 4 and finally only 2 participants scoring the highest as 5.

Table 9 - 3rd Test's statistics

	3rd Test
mean	1.74
median	2.00
mode	2.00
standard dev	1.08



3rd Test Scores all Participants	
Scores	Participants
0	11
1	32
2	36
3	16
4	3
5	2
6	0
7	0

Figure 11 - 3rd Test's scores for all particpatns

Figure 12 summarizes the distribution of participants by their unweighted scores (ranging from 0 to 7) for the 3 tests.

Tests	Number of Participants by Score							
Score	7	6	5	4	3	2	1	0
Test 1	17	24	27	18	12	2	0	0
Test 2	13	15	18	19	14	13	6	2
Test 3	0	0	2	3	16	36	32	11

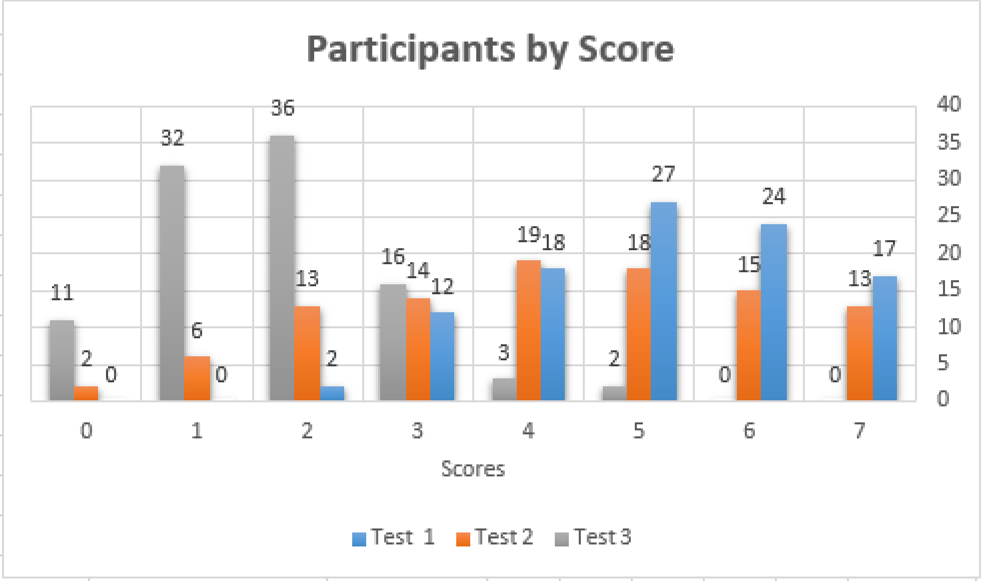


Figure 12 – Participants grouped by their unweighted scores for the three tests

Despite the fact that the scores on each Test is affected by a gradual level of difficulty built into the test, however, the comparisons between the scores of the three Tests at this stage are carried out using the standardized scores not considering the given weight of the difficulty factor in the comparison. The reason for not using the weighted scores is because the goal of these comparisons is to analyse and understand the distribution of the standardized scores over the 100 participants in each test as opposed to the hypotheses tests (Section 5.12) whose goal is to measure and compare the level of awareness improvements reflected by the participants' scores, hence the weighted scores are used there instead. Next, participants' demographics are discussed.

5.6 Participants' Demographics

From this point onward, participants' demographics refer to the 100 Participants who completed the 5 phases of their participation only whose demographical data are collected as part of answering the 1st Test. This data includes different attributes and demographics of the participants such as age group, education level and major, profession, and number of years spent in online study. In the following sections, these participants' attributes and demographics are discussed in detail.

5.7 Participants' Age Groups

As the only criterion for selecting participants for this research was that they are studying online, other attributes such as gender was not part of the selection criteria. However, participants were required to answer the age question by selecting one of 7 different age groups. Each age group represents a 5 year span ranging from 20 to 51+ years as illustrated in Table 14. The analysis of these age groups revealed that the majority of participants (74%), are in the 20-25 age group or alternately said, all 60 participants in the Face-to-Face Group are in the 20-25 age group whereas (14/40) participants from the Online Participant group (35%) are in the 20-25 age group. The next 10% of participants are all from the Online Participant group and are aged between 26 to 30 years old. All the other remaining 16% of participants are all also from the Online Participant group and are distributed over the other 5 age groups. Table 10, Figure 13 and Figure 14 show the age groups break-down of the Face-to-Face Participants and the Online Participant groups and all participants respectively.

Table 10 - Age Groups Break-down for the Face-to-Face and Online particpants

Participants' Age Group	Face to Face Group	Online Group	Grand Total
51		1	1
20 - 25	60	14	74
26 - 30		10	10
31 - 35		8	8
36 - 40		3	3
41 - 45		2	2
46 - 50		2	2
Grand Total	**60**	**40**	**100**

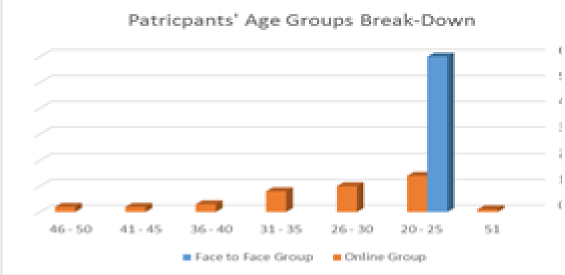


Figure 13 - Age Groups Break-down for the Face-to-Face and Online particpatns

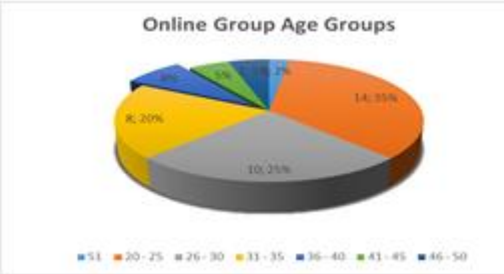


Figure 14 - Age Groups Break-down for all participants combined

5.8 Participants' Education Levels and Major Field Studies

Participants were requested to select one of 4 different education levels ranging from High School or below, through Bachelor and Masters' Degrees to PhD to describe their education levels. Eventually, analysis of participants' education levels show that the majority of the Face-to-Face participant group (57/60) are in the High School level or below, while the majority of the Online Participant group (32/40) have a bachelor degree. This explains the reason why all the Face-to-Face participant group are younger (in the 20-25 age group) than their counterparts in the Online Participant group who are distributed over the other different 7 age groups. Table 11, Figure 15 and Figure 16 illustrate the participants' educational levels break-down.

Table 11 – Educational Levels break-down for the Face-to-Face and Online participants

Participants Education Level	Face to Face Group	Online Group	Grand Total
Bachelor Degree	2	32	34
High School or Below	57	2	59
Master's Degree		4	4
PhD	1	2	3
Grand Total	**60**	**40**	**100**

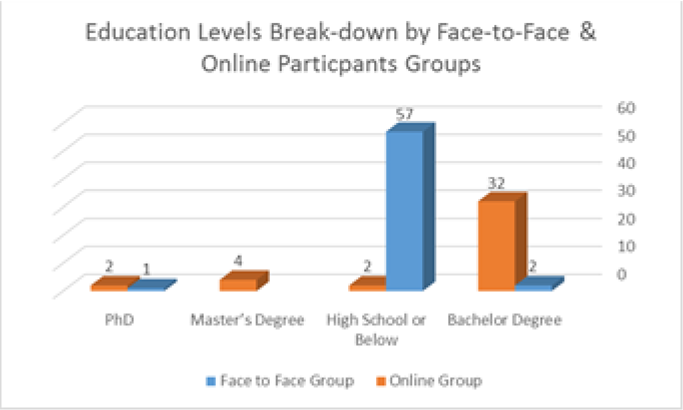


Figure 15 - Educational Levels break-down for the Face-to-Face and Online participants

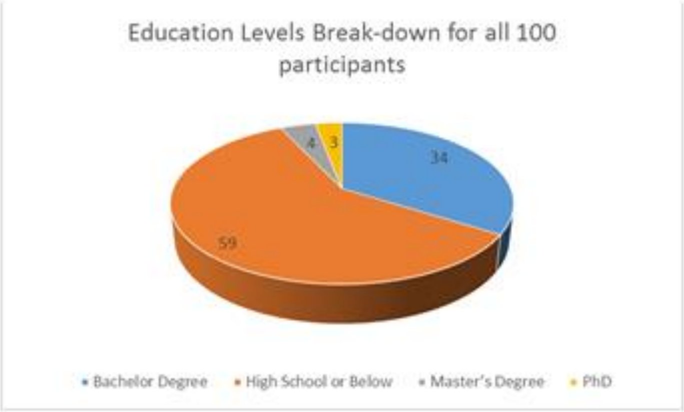


Figure 16 - Educational Levels break-down for all participants combined

Dissecting participants' education levels into majors as in Table 12 shows that participants study for 10 different disciplines and fields such as IT, education, Engineering, Medicine, Oil and Gas, Business Administration and Marketing, Fine Arts and Languages, Human Resource Management, Religious Studies and Science. The top 3 disciplines are High School with the highest majority of participants (40%), followed by the Oil and Gas field which has 17% of participants and finally IT with 13% of the participants. The other 30% participants are distributed over the remaining 7 disciplines as shown in Table 12 and Figure 17.

Table 12 - Participants' Educational Levels by major

Participants Major	Bachelor Degree	High School or Below	Master's Degree	PhD	Grand Total
Business Administration/ Marketing	7		2		9
Education			2		2
Engineering	4			2	6
Fine Arts & Languages	4	2		1	7
HR Management	2				2
IT	12	1			13
Medicine	3				3
Oil and gas		17			17
Religion Studies	1				1
Science	1	39			40
Grand Total	**34**	**59**	**4**	**3**	**100**

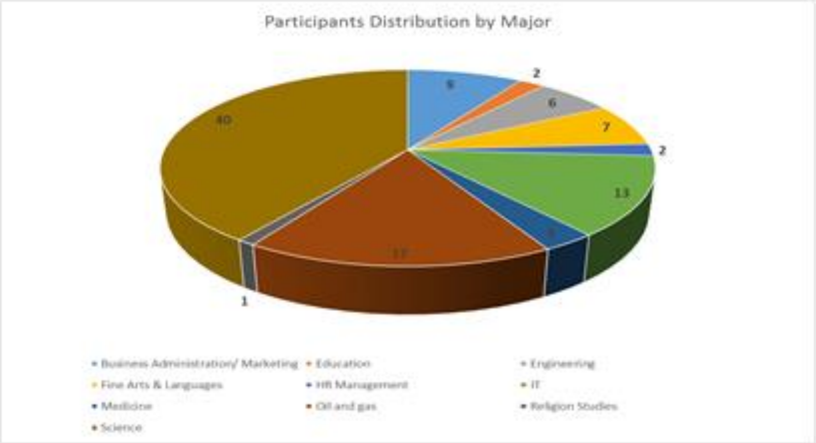


Figure 17 - Participants' Educational Levels by major

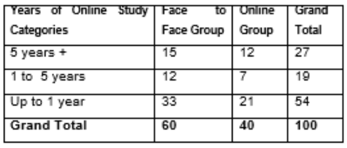
5.9 Participants' Jobs and Number of Years in Online Study

Participants' years of online study is the only criterion for accepting them to participate in this project. However, participants have spent different number of years in online study. Participants were requested to select one of three categories ranging from below 1 to above 5 years; based on the period of online study they have spent. Table 14 shows these 3 categories and the break-down of participants by their jobs. The statistics below show that the majority of participants (54%) have spent up to one year studying online and most of these participants are trainees with High School level qualification.

The second highest category is 27% of participants who spent in online study up to and beyond 5 years. Again, the majority of these participants (14% of all participants; equivalently 52% of the '5 Years +' category participants) are trainees as well. Finally, the lowest category in number of participants is those who are in the middle group spending over 1 year and below 5 years. 19% of participants are in this category (1 to 5 Years) with the majority (12% of all participants; equivalently 63% of the '1 to 5 Years' category participants) being trainees as well. This makes the Trainee participants represent 57% of all participants; which explains the correspondence between this number and the number of participants with High School level qualification (59%) previously explained on participants' education levels.

The other 43% participants are represented by 8 different job categories such as Manager, Middle Management, Regular Employees, Senior Management, Team Leaders, Vice President, Unemployed and Others. Most of these 43 participants (23% of all participants) have spent up to 1 year in online study, 13% have spent 5 or more years and finally 7% have spent from 1 to 5 years in their online study. Table 13 and Figure 18 show these statistics in detail.

Table 13 - Participatns by number of years in Online Study



Years of Online Study Categories	Face to Face Group	Online Group	Grand Total
5 years +	15	12	27
1 to 5 years	12	7	19
Up to 1 year	33	21	54
Grand Total	**60**	**40**	**100**

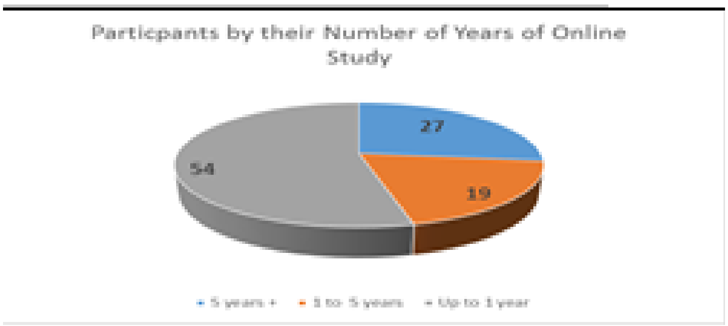
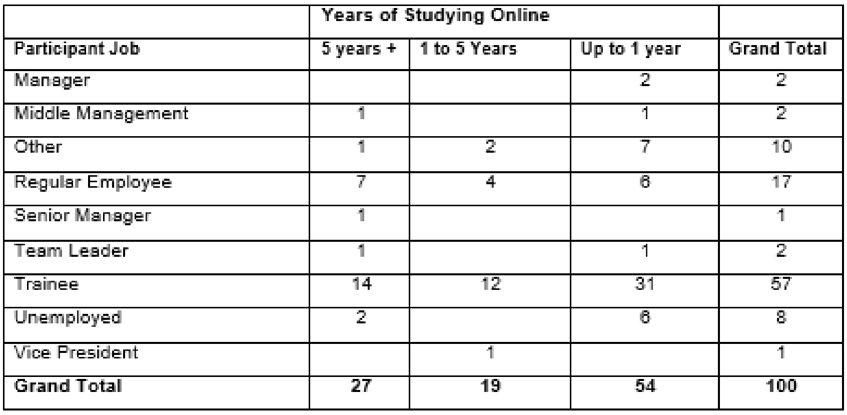


Figure 18 -Participants by number of years in Online Study

Table 14 - Participants' Years of online study by Job



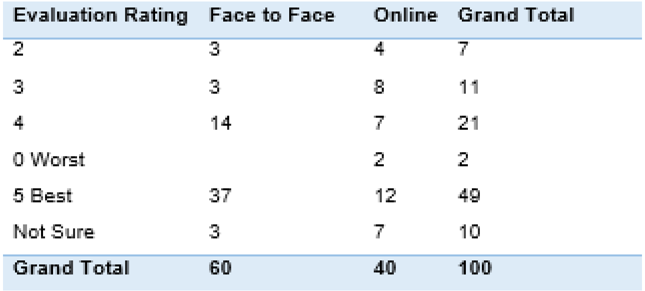
	Years of Studying Online			
Participant Job	5 years +	1 to 5 Years	Up to 1 year	Grand Total
Manager			2	2
Middle Management	1		1	2
Other	1	2	7	10
Regular Employee	7	4	6	17
Senior Manager	1			1
Team Leader	1		1	2
Trainee	14	12	31	57
Unemployed	2		6	8
Vice President		1		1
Grand Total	**27**	**19**	**54**	**100**

5.10 Participants' Evaluation of ISS Role

Last, but not least, Participants' evaluation of Information Security Support (ISS) role at their workplace or school was obtained from participants in order to evaluate the ISS role from the participants' standpoint in spreading awareness about Phishing and study the impact of ISS role on Phishing awareness raising. Participants were requested to choose a category on a 6 point Likert scale (Cohen, Manion and Morrison, 2011) to evaluate the ISS role. The Likert scale represents 6 categories ranging from 'Worst' to 'Best' where 'Worst' is the lowest rating and 'Best' is the highest. The Likert scale contains one more option for the 'Not Sure' category to represent neutrality. 'Excellent', 'Very Good' and 'Good' represent the middle categories on the Likert scale between 'Best' and 'Worst'.

Table 15 and Figure 19 indicate almost half of the participants are highly satisfied with their ISS. Thus, 49% of the participants have rated their ISS as 'Best', followed by 21% as 'Excellent', 11% as 'Very Good', 7% as 'Good' and 2% as 'Worst' respectively while 10% of participants opted to be neutral by selecting the 'Not Sure' option. 37/49 participants who rated their ISS as 'Best' are aged between 20 to 25 years and are part of the Face-to-Face group except for only 2 who are from the Online Group. The majority of the 49 participants (36/49 participants) are in the high school education level too. Next, the tests' scores analysis is discussed.

Table 15 - Participants' evaluations of ISS role in their organizations



Evaluation Rating	Face to Face	Online	Grand Total
2	3	4	7
3	3	8	11
4	14	7	21
0 Worst		2	2
5 Best	37	12	49
Not Sure	3	7	10
Grand Total	60	40	100

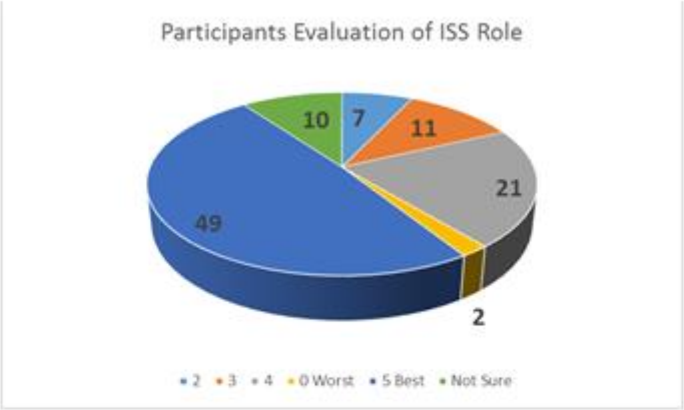


Figure 19- Participants' evaluations of ISS role in their organizations

5.11 Tests' Scores Analysis

The SAMFP model (see Chapter 3 for more details) was designed and applied to raise awareness for a 100 participants who went through a series of 3 tests intervened by 2 awareness sessions successively. The outcomes of this experiment were the collected participants' scores from the three tests. To be able to understand and interpret the collected scores in accordance with the expected goals of the research and the proposed SAMFP model, a set of null hypothesis (Cohen, Manion and Morrison, 2011) were established to help in the analysis of these scores. The analysis of the scores depends on statistically testing the scores for significance and verify whether the objectives and expected learning outcomes of SAMFP had materialized. Thus, the collected scores from the three tests are compared and contrasted with each other to identify variances. These variances are then analysed using a group of statistical tests to find out whether these variances are statistically significant. The next section contains the details of the hypotheses and the analysis procedure to statistically test each hypothesis.

5.12 Tests' Scores Analysis Procedure

The 100 participants are divided into two groups based on the setup of the awareness sessions they attended whether online (40 participants) using Google Hangouts (Isaacson, 2013) or Face-to-Face (60 participants) using a classroom based setup. The scores from the three tests are categorized into the following 9 datasets:

1. Dataset of the 1st Test's scores which contains all 100 participants including Face-to-Face and Online groups
2. Dataset of the 1st Test's scores which contains the Face-to-Face group only
3. Dataset of the 1st Test's scores which contains the Online group only
4. Dataset of the 2nd Test's scores which contains all 100 participants including Face-to-Face and Online groups
5. Dataset of the 2nd Test's scores which contains the Face-to-Face group only
6. Dataset of the 2nd Test's scores which contains the Online group only
7. Dataset of the 3rd Test's scores which contains all 100 participants including Face-to-Face and Online groups
8. Dataset of the 3rd Test's scores which contains the Face-to-Face group only
9. Dataset of the 3rd Test's scores which contains the Online group only

The separation of participants into these two different groups gives an opportunity to study and analyse the effect of Space which is a variable of the SAMFP model on the learning outcomes. The Space variable is demonstrated by the online setup opposed to the Face-to-Face classroom setup through which the awareness sessions of SAMFP are delivered. The other variable that may affect the outcomes of the SAMFP model is Time. The Time variable is represented by the average time span over which each participant completed the five stages of SAMFP. These time spans are not the same for all participants as they tend to complete the course consuming different periods of time. The representation of the Time variable and its effect on the outcomes of SAMFP will be discussed further in the Results Discussion Chapter.

The 9 datasets are statistically compared and contrasted for any significant variances between the tests' scores. The results from all the statistical tests are correlated to find any potential relationships that could influence and impact the scores of the participants. Nonetheless, some of the null hypotheses are stated to test the impact of some of the demographics of participants on the scores such as the participants' age, education, number of years of online study and the role of ISS. Hence, the null hypotheses are divided into three categories according to the scope of analysis:

1. Hypotheses for an individual group's scores across the three tests e.g. comparing the Online group's scores on the 1st Test with the 2nd Test, the 2nd Test with the 3rd Test and the 3rd Test with the 1st Test.
2. Hypotheses for two independent groups' scores within each test e.g. comparing the Online Group's scores with the Face-to-Face group's scores on the 1st Test, then on the 2nd Test and finally on the 3rd Test.
3. Hypotheses for analysis of participants' demographics' influence on the scores of each test e.g. the impact of participants' age on the scores on the 1st Test, 2nd Test and 3rd Test.

In order to achieve the objectives of the research and be able to answer the research questions of whether SAMFP has been able to help the participants' improve and take their awareness about Phishing behavioural factors to the next levels, 18 null hypotheses are stated.

5.13 Hypotheses

The first set of hypotheses consists of 3 Null and alternate hypotheses pairs aimed at testing whether there is a significant difference between the Face-to-Face and Online participant groups in gaining knowledge and improving awareness about Phishing behavioural factors. These hypotheses are statistically tested by comparing the mean score of the participants in one group with the mean score of the participants in the other group to find out if the two compared mean scores are significantly different. If the test result is significant, then the Null hypothesis will be rejected, otherwise supported and the Alternate hypothesis will be rejected. Each null/alternate hypotheses pair is designed to test a level of participant awareness according to the Endsley (2015) Situation Awareness levels: 'Perception', 'Comprehension' and 'Projection'.

The second set of Null/Alternate hypotheses also consists of three hypothesis pairs. These 3 pairs aim to test whether there is a significant difference in the participants' scores among the three tests. However, this time the scores compared between the three tests are for the participants in the same group e.g. the Online participant group or the Face-to-Face group or all participants (combining online and Face-to-Face) as one group. . If the result of the test shows a significant difference between the scores, the Alternate hypothesis is accepted and the Null hypothesis is rejected. This means that the respective participant groups (Online, Face-to-Face or all) either improved their awareness about Phishing after attending the two awareness sessions, or maintained the same/lower awareness level on the higher tests taking into account the fact that the higher tests are of increasing difficulty and weighted. For example, participants' scores on the 2nd and 3rd Tests' scores are lower than their scores on the 1st Test.

This set of Null/Alternate hypothesis pairs do not specify on which test, participants performed significantly better and eventually gained higher levels of awareness. For this reason, the third set of Null/Alternate hypothesis pairs are designed to compare individual pairs of datasets to identify which datasets have significantly made the difference among all compared pairs of datasets. Therefore, nine (9) Null/Alternate pairs of hypotheses are stated to test whether the mean scores on the 2nd Test is significantly higher than the mean scores on the 1st Test considering that the participants attended the 1st awareness session before the 2nd Test. Similarly, they test whether the mean scores on the 3rd Test is significantly higher than the mean scores on the 2nd Test and 1st Test respectively considering that participants had already attended the 1st and 2nd awareness sessions; one before the 2nd Test and the other one before the 3rd Test. It is worth noting that the level of difficulty on every test also varies and thus is taken into consideration when comparing the mean scores between the three tests. This is accomplished by giving each test a specific weight that meets its level of difficulty. More on this has already been discussed in Chapter 4.

The statistical tests chosen to compare these 9 Null hypotheses are two tailed t-tests (Zimmerman, 2017) aimed at verifying whether the participant awareness levels about Phishing had improved gradually after taking the awareness sessions. Each pair of hypotheses test the awareness improvements for the Online participant group, the Face-to-Face participant group and all participants combined against the three assessment tests.

In summary, there are 3 Null/Alternate pairs of hypotheses comparing the mean scores of the 2nd Test with the 1st Test for all participants combined, Online participant group and Face-to-Face participant group to verify the Perception level. In the same way, another set of 3 hypothesis pairs compare the Mean Scores of the 3rd Test with the 2nd Test for all participants combined, Online participant group and Face-to-Face participant group to verify if their awareness reached the ‘Comprehension Level’. Finally, another set of 3 hypothesis pairs comparing the mean scores of the 3rd Test with the 1st Test’s means scores for all participants combined, Online participant group and Face-to-Face participant group to verify if their awareness reached the ‘Projection’ Level.

If the result of any of these nine (9) tests does not significantly support the Null hypothesis and hence supports the Alternate hypothesis, this means two possibilities to be checked. First, the mean scores on the compared Test e.g. the 2nd Test is significantly higher than the mean scores on the compared with Test e.g. the 1st Test, indicating that the awareness session given to the participants after the 1st Test and before the 2nd Test had significantly contributed to the improvement of participants' level of awareness. The second possibility is that the mean scores of the 1st Test is significantly higher than that of the 2nd Test or the 3rd Test or the mean scores on the 2nd Test is significantly higher than the mean scores on the 3rd Test indicating that the awareness sessions given to the participants did not adequately contribute to improve the participants' level of awareness. The theory suggests that more iterations of awareness sessions may still be needed for the expected awareness improvements to materialize. In either case (possibility), the resulting outcomes are in line with the theory of the proposed situation awareness conceptual framework SAMFP model.

The last set of Null/Alternate pairs of hypotheses to test are aimed to assess the effect of participants' demographical attributes on the scores and awareness achievements and to also use these participants' attributes such as age groups, education level and number of years of online study to see if that can help predict the participants' future scores on the 1st, 2nd and 3rd Tests through the use of Regression tests (Draper and Smith, 196).

See Appendix B for a list of all the 18 hypotheses. Now after the 18 Null/Alternate hypotheses pairs have been defined, the results of each tested hypothesis are explained next. Please see Appendix A for details on the design of the statistical testing process conducted on each hypothesis.

5.14 Hypotheses Testing

The 18 hypotheses will be statistically tested, analysed, compared and correlated and thus will provide the statistical information required to make meaningful conclusions about the assumptions made about the data in response to the research questions. Following the steps of the testing process explained in Appendix A and guided by the results of the normality and variance equality tests conducted on the data as part of the testing process, each hypothesis will be tested by the proper statistical tests. The results will help determine whether to support the null or the alternate hypothesis. After all the 18 hypotheses are tested and the results are analysed, more discussion of the results will take place in the next chapter 'Results Discussion'. Now the testing of the first 3 hypotheses null/alternate pairs is discussed.

5.15 Score Variance between Online and Face-to-Face Participant Groups within each Assessment Test

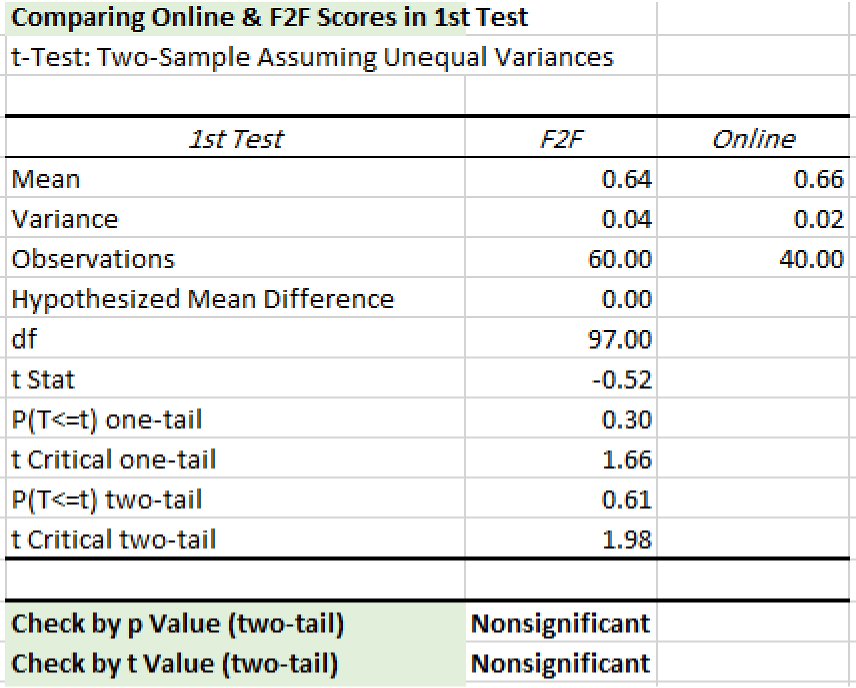
The first three null hypotheses make an assumption that there is no significant difference in the level of awareness gain between the Online and the Face-to-Face participants on the 1st, 2nd and 3rd Tests. In order to test if these hypothesized assumptions are true, the means of the weighted scores of the two groups of participants are statistically compared using an Independent Two Sample t-Test (Cohen, Manion and Morrison, 2011) after having verified that both datasets meet the normality and variance equality assumptions required by the Independent Samples t-Test. The Independent Two Sample t-Test is conducted three times comparing the two groups' scores once for the 1st Test, then the 2nd Test and finally for the 3rd Test: The results of these tests are as follows.

H1: Comparing Datasets 2 and 3

Null Hypothesis (H1a): There is no significant difference between the mean scores of the Online Group and the Face-to-Face Group on the 1st Test.

Test Results: Since the Face-to-Face and Online Groups' scores on the 1st Test have been verified to have an acceptable level of normal distribution and the results of the F-Test on the compared datasets indicated unequal variance, a two tailed Two Sample t-test assuming unequal variances is conducted. The test shows that (p (two tailed) = 0.60 > a = 0.05, t statistic = -0.518, df = 97) as in Table 16. In conclusion, the results support the null hypothesis (H1a) that there is no significant difference between the mean scores of the Online Group (M = 0.65. SD= 0.14) and the Face-to-Face Group (M = 0.63. SD= 0.19) on the 1st Test and hence the Alternate hypothesis (H1b) is rejected.

Table 16 - Two Sample t-test Assuming Unequal Variances Comparing Face-to-Face and Online Groups' scores on 1st Test



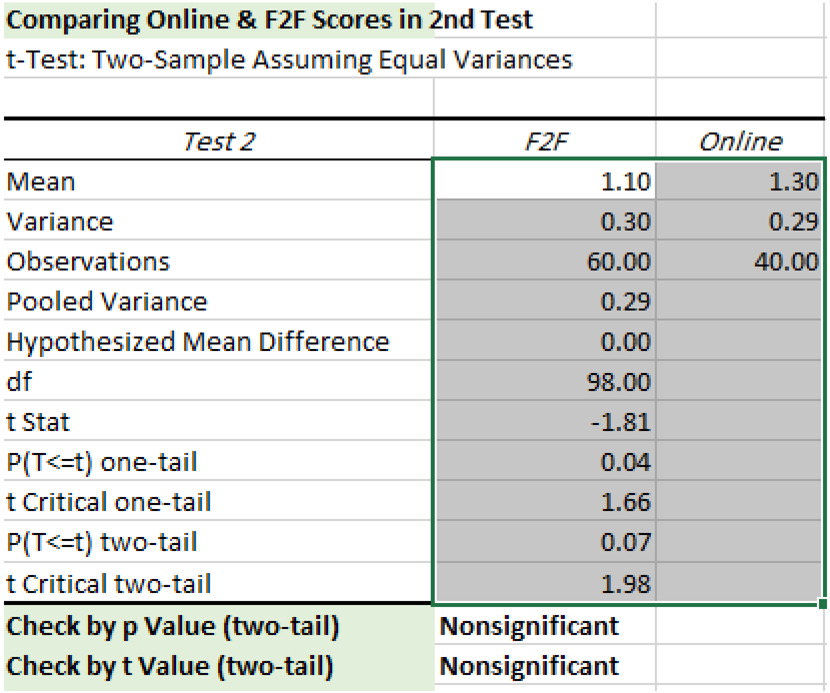
Comparing Online & F2F Scores in 1st Test		
t-Test: Two-Sample Assuming Unequal Variances		
1st Test	*F2F*	*Online*
Mean	0.64	0.66
Variance	0.04	0.02
Observations	60.00	40.00
Hypothesized Mean Difference	0.00	
df	97.00	
t Stat	-0.52	
P(T<=t) one-tail	0.30	
t Critical one-tail	1.66	
P(T<=t) two-tail	0.61	
t Critical two-tail	1.98	
Check by p Value (two-tail)	**Nonsignificant**	
Check by t Value (two-tail)	**Nonsignificant**	

H2: Comparing Datasets 5 and 6

Null Hypothesis (H2a): There is no significant difference between the mean scores of the Online Group and the Face-to-Face Group on the 2nd Test.

Test Results: Since the Face-to-Face and Online Groups' scores on the 2nd Test have been verified to have an acceptable level of normal distribution and equal variances, a two tailed Two Sample t-test assuming equal variances is conducted as shown in Table 17. The test results indicated that there is no significant difference between the mean scores of the F2F Participants Group (M= 1.09. SD= 0.54) and the Online Participants Group (M=1.29, SD= 0.54) on the 2nd Test where (p (two tailed) = 0.07 > a = 0.05, t statistic = -1.80, df = 98). Hence, the results support the Null hypothesis (H2a) and thus the Alternate hypothesis (H2b) is rejected.

Table 17 - Using t-Test to compare Face-to-Face and Online groups' scores on 2nd Test



Comparing Online & F2F Scores in 2nd Test

t-Test: Two-Sample Assuming Equal Variances

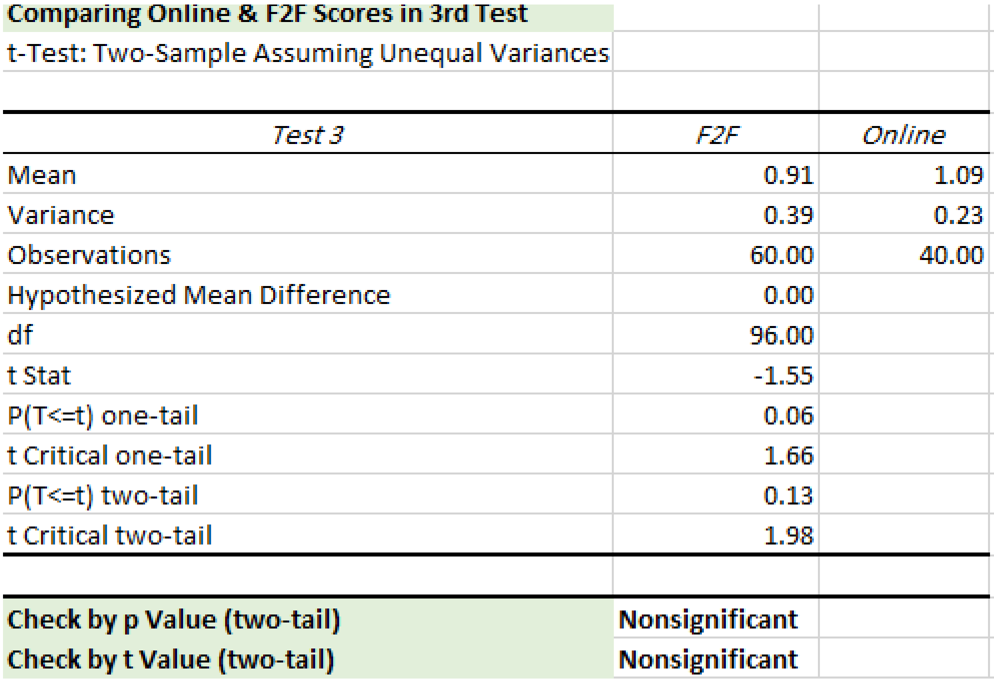
Test 2	*F2F*	*Online*
Mean	1.10	1.30
Variance	0.30	0.29
Observations	60.00	40.00
Pooled Variance	0.29	
Hypothesized Mean Difference	0.00	
df	98.00	
t Stat	-1.81	
P(T<=t) one-tail	0.04	
t Critical one-tail	1.66	
P(T<=t) two-tail	0.07	
t Critical two-tail	1.98	
Check by p Value (two-tail)	**Nonsignificant**	
Check by t Value (two-tail)	**Nonsignificant**	

H3: Comparing Datasets 8 and 9

Null Hypothesis (H3a): There is no significant difference between the mean scores of the Online Group and the Face-to-Face Group on the 3rd Test.

Test Results: Since the Face-to-Face and Online Groups' scores on the 3rd Test have been verified to have an acceptable level of normal distribution and the results of the F-Test on the compared datasets indicated unequal variance, a two tailed Two Sample t-test assuming unequal variances is conducted. The test results indicated that there is no significant difference between the mean scores of the Face-to-Face participant group (M = 0.91. SD= 0.62) and the Online participant group (M = 1.08, SD= 0.48) on the 3rd Test where (p (two tailed) = 0.12 > a = 0.05, t statistic = -1.54, df = 96) as in Table 18. In conclusion, the results of the test conducted support the null hypothesis (H3a) that there is no significant difference between the mean scores of the Online Group and the Face-to-Face Group on the 3rd Test and hence the Alternate hypothesis (H3b) is rejected.

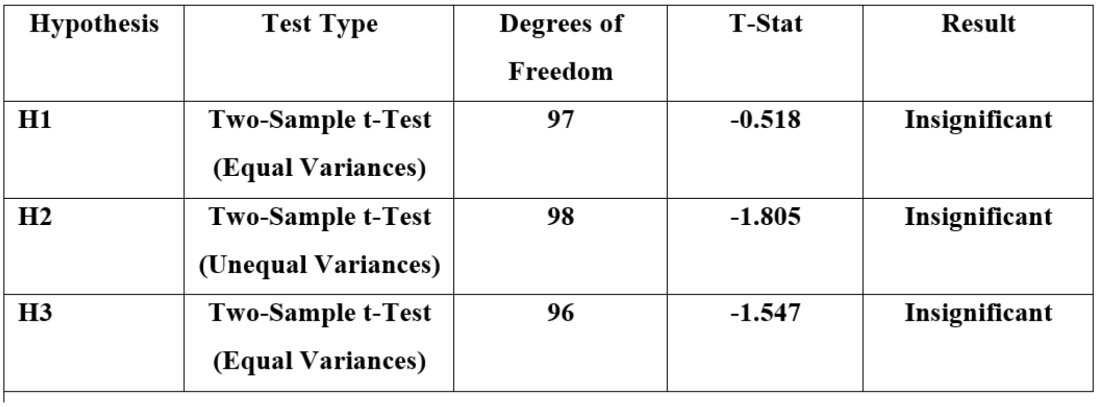
Table 18 - Two Sample t-test assuming unequal variances comparing Face-to-Face and Online Groups' scores on 3rd Test



Comparing Online & F2F Scores in 3rd Test		
t-Test: Two-Sample Assuming Unequal Variances		
Test 3	*F2F*	*Online*
Mean	0.91	1.09
Variance	0.39	0.23
Observations	60.00	40.00
Hypothesized Mean Difference	0.00	
df	96.00	
t Stat	-1.55	
P(T<=t) one-tail	0.06	
t Critical one-tail	1.66	
P(T<=t) two-tail	0.13	
t Critical two-tail	1.98	
Check by p Value (two-tail)	**Nonsignificant**	
Check by t Value (two-tail)	**Nonsignificant**	

In summary as in Table 19 below, the 3 hypotheses (H1, H2 and H3) were tested to determine if there was a statistically significant difference between the Face-to-Face participants and the Online participants in acquired awareness as reflected by their scores on the 1st, 2nd and 3rd Tests respectively. The tests' results show that there was no significance difference in the level of awareness gained between the two groups on the 1st, 2nd and the 3rd Tests. It is noted that H1, H2 and H3 have different degrees of freedom despite the compared samples are the same in number. This is due to the differences in variances and standard deviations of the two samples. For example, H2 has equal variances where degrees of freedom are always rendered the same following the formula ((n1 + n2) – 2). On the other hand, H1 and H3 have unequal variances and therefore, the formula tends to give different results depending on how large or small the differences between the standard deviations of the two samples. MS Excel data analysis tool was used to generate these tests where it tends to round off the degrees of freedom to the next integer in case of unequal variances (Real.Statistics.com, 2017).

Table 19 - Results Summary of H1, H2 and H3



Hypothesis	Test Type	Degrees of Freedom	T-Stat	Result
H1	**Two-Sample t-Test (Equal Variances)**	**97**	**-0.518**	**Insignificant**
H2	**Two-Sample t-Test (Unequal Variances)**	**98**	**-1.805**	**Insignificant**
H3	**Two-Sample t-Test (Equal Variances)**	**96**	**-1.547**	**Insignificant**

5.16 Score Variance between Assessment Tests for each Participant Group

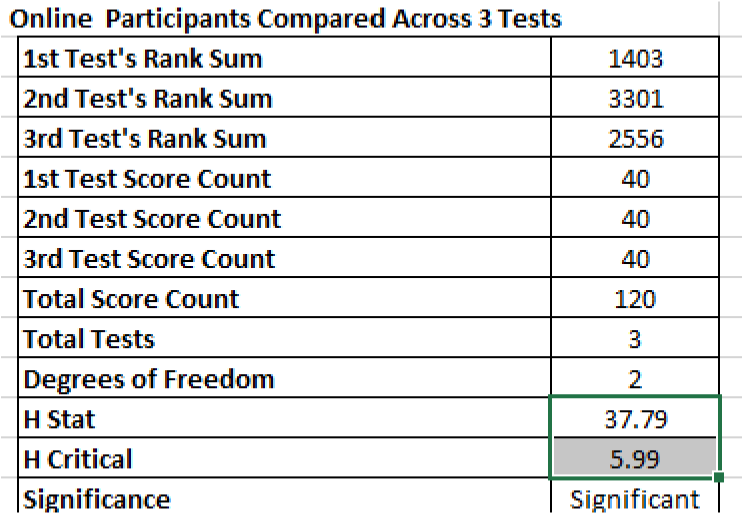
The next 4th, 5th and 6th hypotheses test if there were significant awareness differences by the same group of participants between the three different Tests:

H4: Comparing Datasets 3, 6 and 9

Null Hypothesis (H4a): There is no significant difference between the mean scores acquired on the 1st, 2nd and 3rd Tests by the Online participant group.

Test Results: Analysis of Variance (ANOVA) is the appropriate statistical test to compare more than two datasets (Cohen, Manion and Morrison, 2011). In this case, three datasets are compared by ANOVA. These three datasets represent the scores of the three tests respectively. Having already tested these three datasets for variance equality, the results indicated that the third dataset representing the 3rd Test's scores compared to the 1st Test's scores, had unequal variances which makes the whole idea of using ANOVA to compare the three datasets infeasible. Therefore, the non-parametric test "Kruskal-Wallis H" will be used as an alternative test to compare the three datasets. Thus, Kruskal-Wallis H Test is conducted to compare the scores of the Online participant Group between the 1st Test and 2nd Test, between the 2nd Test and the 3rd Test and finally between the 3rd Test and 1st Test. The results of the Kruskal-Wallis H test indicates that there is a significant difference in the median scores between the 1st, 2nd and 3rd tests' scores for the Online participant group at the alpha level 0.05 with H Statistic = 37.78 > H Critical = 5.991 as shown in Table 20. Therefore, the Null hypothesis (H4a) is not supported, hence the Alternate (H4b) is. This indicates that there is a significant difference in the awareness levels between the three Tests' scores for the Online participant group. However, to determine on which test the significant difference resides, further individual statistical tests are conducted as part of the remaining hypotheses that follow where the effect size of any significant variance found is also measured.

Table 20 - Kruskal-Wallis H test comparing Online participant group' scores across the 3 tests



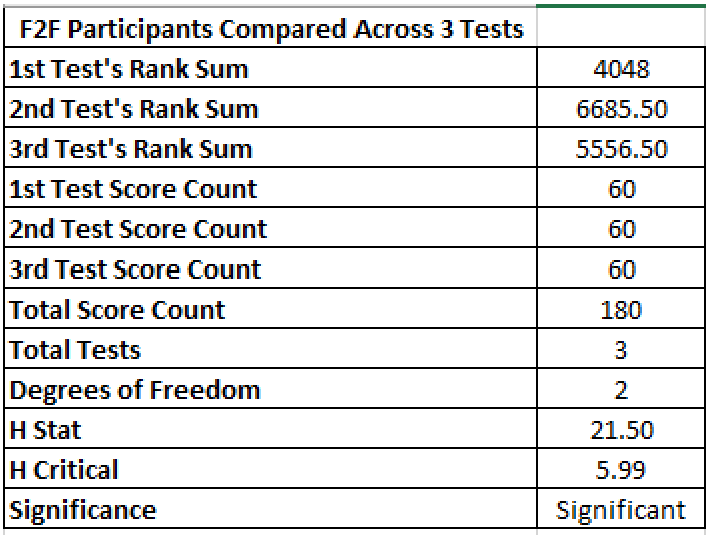
Online Participants Compared Across 3 Tests	
1st Test's Rank Sum	1403
2nd Test's Rank Sum	3301
3rd Test's Rank Sum	2556
1st Test Score Count	40
2nd Test Score Count	40
3rd Test Score Count	40
Total Score Count	120
Total Tests	3
Degrees of Freedom	2
H Stat	37.79
H Critical	5.99
Significance	Significant

H5: Comparing Datasets 2, 5 and 8

Null Hypothesis (H5a): There is no significant difference between the mean scores acquired on the 1st, 2nd and 3rd Tests by the Face-to-Face participant group.

Test Results: Similar to the previous hypothesis (H4a), three datasets are compared using Kruskal-Wallis H test, however hypothesis (H5a) tests the Face-to-Face participant group' scores across the three Tests. Thus, Kruskal-Wallis H Test was conducted to compare the scores of the Face-to-Face participant group between the 1st Test and 2nd Test, between the 2nd Test and the 3rd Test and finally between the 3rd Test and 1st Test. The results of the Kruskal-Wallis H test indicate that there is a significant difference in the median scores between the 1st, 2nd and 3rd tests' scores for the Face-to-Face participant group at the alpha level 0.05 with H Statistic = 21.49 > H Critical = 5.991 as shown in Table 21. Hence, the Null hypothesis (H5a) is not supported and therefore the Alternate (H5b) is supported indicating a significant difference in the awareness levels between the three Tests for the Face-to-Face participant group. However, to determine on what test the significant difference has emerged, further individual statistical tests are conducted as part of the remaining hypotheses that follow where the effect size of any significant variance found is also measured.

Table 21 - Kruskal-Wallis H test comparing the Face-to-Face group's scores across the 3 tests



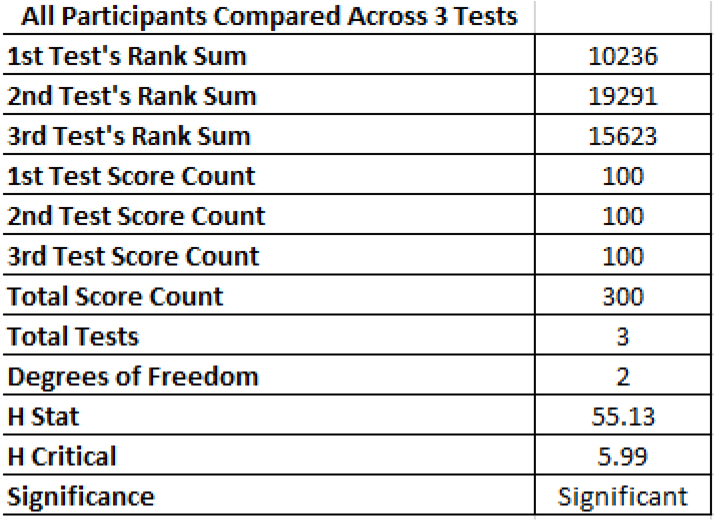
F2F Participants Compared Across 3 Tests	
1st Test's Rank Sum	4048
2nd Test's Rank Sum	6685.50
3rd Test's Rank Sum	5556.50
1st Test Score Count	60
2nd Test Score Count	60
3rd Test Score Count	60
Total Score Count	180
Total Tests	3
Degrees of Freedom	2
H Stat	21.50
H Critical	5.99
Significance	Significant

H6: Comparing Datasets 1, 4 and 7

Null Hypothesis (H6a): There is no significant difference between the mean scores acquired on the 1st, 2nd and 3rd Tests by all participants combining the Online and Face-to-Face Groups participants together.

Test Results: Similar to the previous hypotheses (H4a and H5a), three datasets are compared using Kruskal-Wallis H test, however hypothesis (H6a) compares the scores of all the participants combined from the two participant groups namely the Online and Face-to-Face across the three Tests. Thus, Kruskal-Wallis H Test was conducted indicating a significant difference in the median scores between the 1st, 2nd and 3rd tests' scores for all participants at the alpha level 0.05 with H Statistic = 55.13 > H Critical = 5.991 as shown in Table 22. Hence, the Null hypothesis (H6a) is not supported and therefore the Alternate (H6b) is supported indicating a significant difference in the awareness levels between the three Tests for all participants. However, to locate which datasets caused the significant difference, further individual statistical tests have been conducted as part of the remaining hypotheses that follow where the effect size of any significant variance found is also measured.

Table 22 - Kruskal-Wallis H test comparing the scores of all participants combined across the 3 tests



All Participants Compared Across 3 Tests	
1st Test's Rank Sum	10236
2nd Test's Rank Sum	19291
3rd Test's Rank Sum	15623
1st Test Score Count	100
2nd Test Score Count	100
3rd Test Score Count	100
Total Score Count	300
Total Tests	3
Degrees of Freedom	2
H Stat	55.13
H Critical	5.99
Significance	Significant

In summary, the hypotheses H4, H5 and H6 tested whether a significant difference in awareness gains between the participants had been reflected by their scores across the three Tests. The results show a significant difference between the scores in all three tests and thus the 3 Null hypotheses above were not supported as shown in Table 23. The next set of hypothesise will identify where the significant of differences between the three tests' scores lie and to what extent their effect was.

Table 23 - Results Summary of H4, H5 and H6

Hypothesis	Test Type	Degrees of Freedom	H-Stat	Result
H4	Kruskal-Wallis	2	37.788	Significant
H5	Kruskal-Wallis	2	21.499	Significant
H6	Kruskal-Wallis	2	55.134	Significant

5.17 Score Variance between the 1st and 2nd Tests for each Participant Group

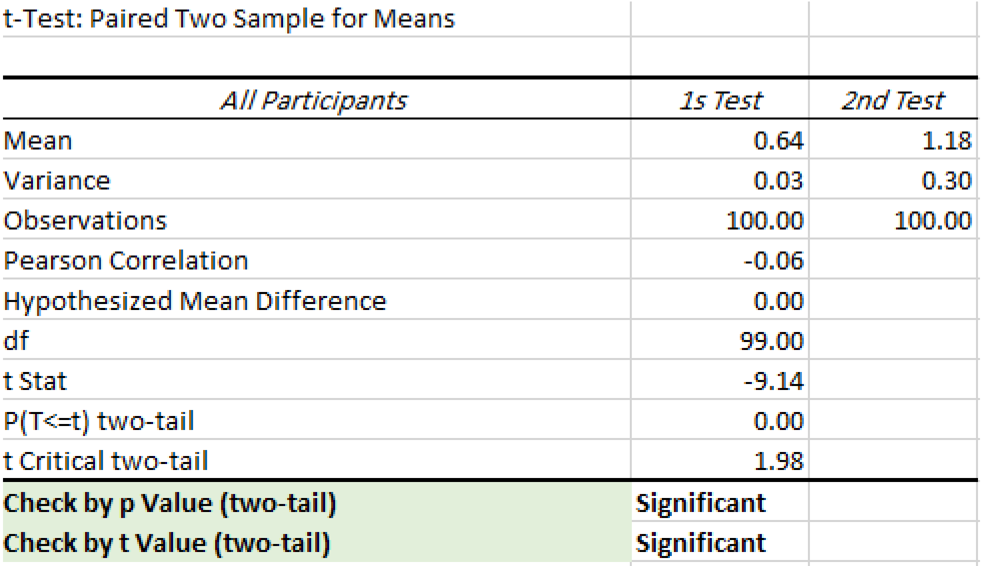
The next 3 sections of hypotheses will test each pair of datasets individually to determine which one of the datasets exactly manifested the significant difference; hence two tailed paired t-tests were used. For example, if the Online participant group had significantly differed in scores from one or two of the three Tests, the individual two tailed paired t-Test will ascertain on which test the Online participant group scored significantly different from the other test; whether higher or lower depending on which test their greater mean score was recorded. Each 3 hypotheses in each section will address a pair of tests to compare the scores of the Online Group, Face-to-Face Group and all participants from both groups combined respectively. As such, the next 3 hypotheses will test the differences in mean scores between the 1st Test and 2nd Test.

H7: Comparing datasets 1 and 3

Null Hypothesis (H7a): There is no statistically significant difference between the mean scores on the 2nd Test and the 1st Test for all participants.

Test Results: A two tailed paired t-test was conducted to compare the mean scores between the 1st Test and 2nd Test for all participants to determine if there was a significant difference between the two tests and which test scores were significantly higher than the other. This is to assess whether the participants had benefited from the awareness session they attended after they had completed the 1st Test and before they had taken the 2nd Test. The participants' scores on the 2nd Test (Experimental Group) are expected to be significantly higher than those on the 1st Test (Control Group) after participants had attended the 1st awareness session. The result of the test shows as in Table 24 a significant difference between the mean scores on the 1st Test (M= 0.64, SD= 0.17) and the 2nd Test (M= 1.18, SD= 0.55) where the t statistic (df = 99) = -9.13 and the P value two-tail = (<0.001) < alpha 0.05. Therefore, it is concluded that the difference between the mean scores on the 1st and 2nd Tests are significantly different where participants scored higher on the 2nd Test compared with the 1st Test considering the higher mean score on the 2nd Test (2nd Test's M= 1.18 > 1st Test's M= 0.64). This conclusion does not support the Null hypothesis (H7a) and hence supports the Alternate Hypothesis (H7b). The effect size of the significance of this mean score difference between the two groups' is 1.49 according to Cohen's d (Trending Sideways, 2017). Cohen's d of 1.49 is considered large size effect according to the range "> 0.8" (Cohen, 1992; Polyu.edu.hk, 2017).

Table 24 - Paired t-test comparing the scores of all participants combined between 1st and 2nd tests



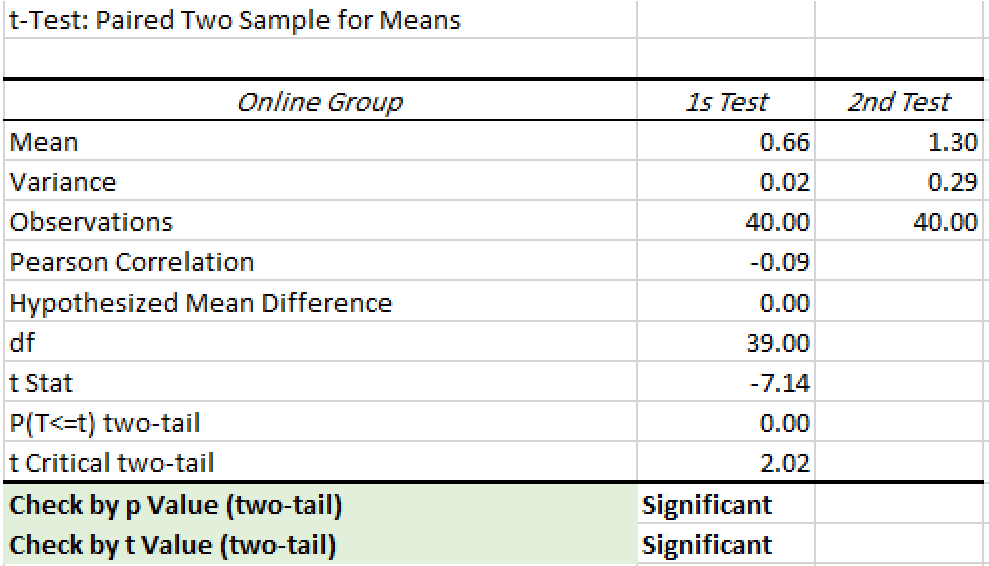
t-Test: Paired Two Sample for Means		
All Participants	*1s Test*	*2nd Test*
Mean	0.64	1.18
Variance	0.03	0.30
Observations	100.00	100.00
Pearson Correlation	-0.06	
Hypothesized Mean Difference	0.00	
df	99.00	
t Stat	-9.14	
P(T<=t) two-tail	0.00	
t Critical two-tail	1.98	
Check by p Value (two-tail)	**Significant**	
Check by t Value (two-tail)	**Significant**	

H8: Comparing datasets 3 and 6

Null Hypothesis (H8a): There is no statistically significant difference between the mean scores on the 2nd Test and the 1st Test for the Online participant group.

Test Results: To test hypothesis (H8a), a two tailed paired t-test is conducted to compare the mean scores between the 1st Test and 2nd Test for the Online participant group to determine if there was a significant difference between the two tests and which test scores were significantly higher than the other. This is to assess whether the Online participant group had benefited from the awareness session they attended after they had completed the 1st Test. The participants' scores on the 2nd Test (Experimental Group) are expected to be significantly higher than those on the 1st Test (Control Group) after attending the 1st awareness session. Similar to hypothesis (H7a), the result of the test shows as in Table 25 a significant difference between the mean scores on the 1st Test (M= 0.66, SD= 0.14) and the 2nd Test (M= 1.30, SD= 0.54) where the t statistic (df = 39) = -7.14 and the P value two-tail = (<0.001) < alpha 0.05. Therefore, it is concluded that the difference between the mean scores on the 1st and 2nd Tests are significantly different where the Online participant group scored higher on the 2nd Test compared with the 1st Test considering the higher mean on the 2nd Test (2nd Test's M= 1.30 > 1st Test's M= 0.66). This conclusion does not support the Null hypothesis (H8a) and hence supports the Alternate Hypothesis (h8b). The effect size of the significance of this mean score difference between the two groups' is 1.89 according to Cohen's d (Trending Sideways, 2017). Cohen's d of 1.89 is considered large size effect according to the range "> 0.8" (Cohen, 1992; Polyu.edu.hk, 2017).

Table 25 - Paired t-test comparing the scores of Online Group between 1st and 2nd test



t-Test: Paired Two Sample for Means

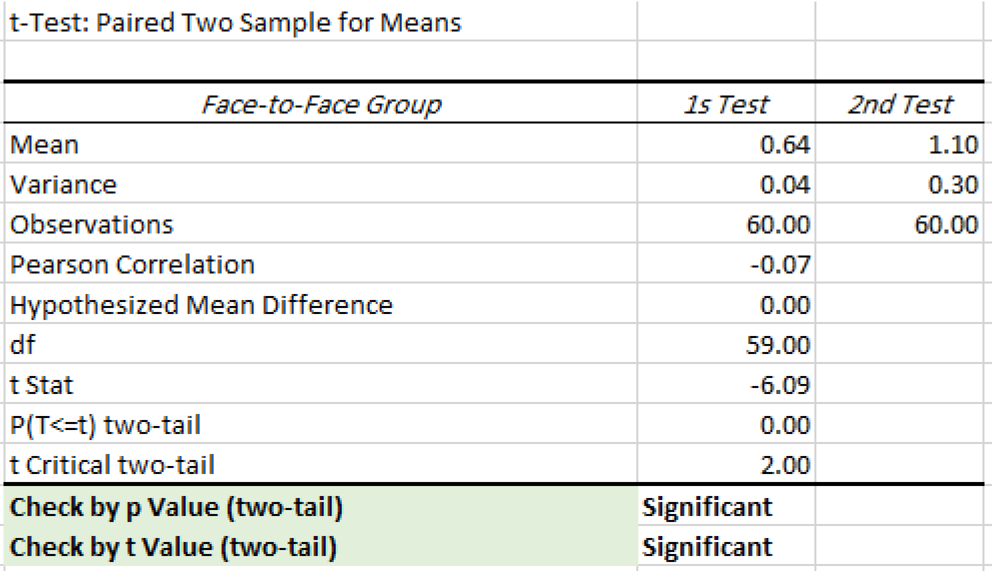
Online Group	*1s Test*	*2nd Test*
Mean	0.66	1.30
Variance	0.02	0.29
Observations	40.00	40.00
Pearson Correlation	-0.09	
Hypothesized Mean Difference	0.00	
df	39.00	
t Stat	-7.14	
P(T<=t) two-tail	0.00	
t Critical two-tail	2.02	
Check by p Value (two-tail)	**Significant**	
Check by t Value (two-tail)	**Significant**	

H9: Comparing datasets 2 and 5

Null Hypothesis (H9a): There is no statistically significant difference between the mean scores on the 2nd Test and the 1st Test for the Face-to-Face participant group.

Test Results: Similar to the Online Group, the Face-to-Face participant group's scores are compared through hypothesis (H9a). Thus, a two tailed paired t-test was conducted to compare the mean scores between the 1st Test and 2nd Test for the Face-to-Face participant group to determine if there was a significant difference between the two tests and on which test, scores were significantly higher than the other. This is to assess whether the Face-to-Face participant group had benefited from the awareness session they attended after they had completed the 1st Test. The participants' scores on the 2nd Test (Experimental Group) are expected to be significantly higher than those on the 1st Test (Control Group) after attending the 1st awareness session. Similar to hypothesis (H8a), the result of the test shows as in Table 26 a significant difference between the mean scores on the 1st Test (M= 0.64, SD= 0.19) and the 2nd Test (M= 1.10, SD= 0.54) where the t statistic (df = 59) = -6.09 and the P value two-tail = (<0.001) < alpha 0.05. Therefore, it is concluded that the difference between the mean scores on the 1st and 2nd Tests are significantly different where the Face-to-Face participant group scored higher on the 2nd Test compared to the 1st Test considering the higher mean on the 2nd Test (2nd Test's M= 1.10 > 1st Test's M= 0.64). This conclusion does not support the Null hypothesis (H9a) and hence supports the Alternate Hypothesis (H9b). The effect size of the significance of this mean score difference between the two groups' is 1.26 according to Cohen's d (Trending Sideways, 2017). Cohen's d of 1.26 is considered large size effect according to the range "> 0.8" (Cohen, 1992; Polyu.edu.hk, 2017).

Table 26 - Paired t-test comparing the scores of Face-to-Face Group between 1st and 2nd test



t-Test: Paired Two Sample for Means		
Face-to-Face Group	*1s Test*	*2nd Test*
Mean	0.64	1.10
Variance	0.04	0.30
Observations	60.00	60.00
Pearson Correlation	-0.07	
Hypothesized Mean Difference	0.00	
df	59.00	
t Stat	-6.09	
P(T<=t) two-tail	0.00	
t Critical two-tail	2.00	
Check by p Value (two-tail)	**Significant**	
Check by t Value (two-tail)	**Significant**	

5.18 Score Variance between the 2nd and 3rd Tests for each Participant Group

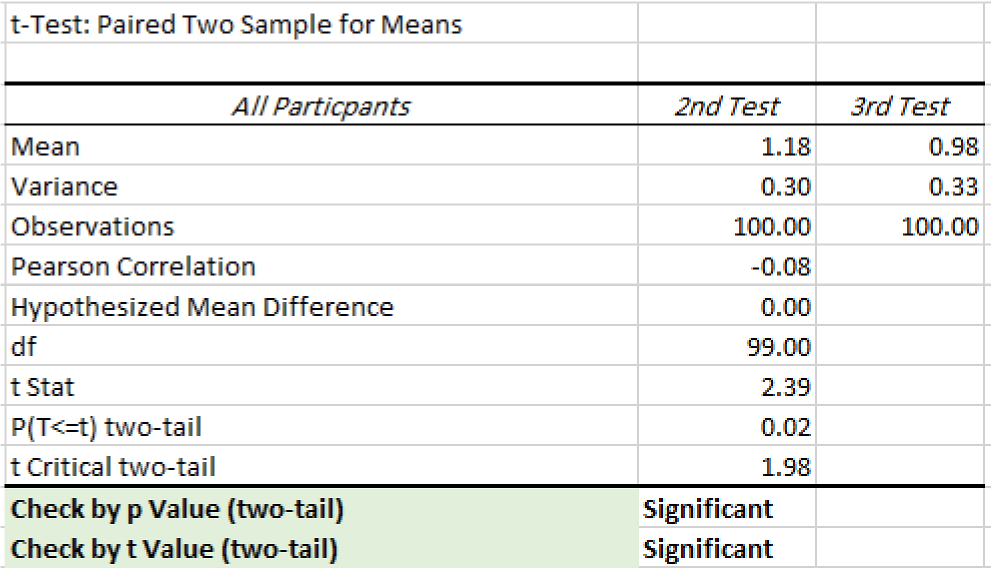
The next 3 hypotheses will assess the differences in mean scores between the 2nd and 3rd Tests in the same pattern as the previous 3 hypotheses above measuring the effect of attending the 2nd awareness session on the participants' level of awareness before and after taking the 3rd test.

H10: Comparing datasets 7 and 4

Null Hypothesis (H10a): There is no statistically significant difference between the mean scores on the 3rd Test and the 2nd Test for all participants.

Test Results: A two tailed paired t-test was conducted to compare the mean scores between the 2nd Test and 3rd Test for all participants to determine if there was a significant difference between the two tests and on which test, scores were significantly higher than the other. This is to assess whether the participants had benefited from the awareness session they attended after they had completed the 2nd Test and before they had taken the 3rd Test. The participants' scores on the 3rd Test (Experimental Group) are expected to be significantly higher than those on the 2nd Test (Control Group) after attending the 2nd awareness session. The result of the test shows as in Table 27 a significant difference between the means scores on the 2nd Test (M= 1.18, SD= 0.55) and the 3rd Test (M= 0.98, SD= 0.57) where the t statistic (df = 99) = 2.39 and the P value two-tail = 0.02 < alpha 0.05. Therefore, it is concluded that the difference between the mean scores on the 2nd and 3rd Tests are significantly different where participants scored higher on the 2nd Test than on the 3rd Test considering the negative effect size explained below. This conclusion does not support the Null hypothesis (H10a) and hence supports the Alternate Hypothesis (H10b). However, the significant difference in the mean scores is in the opposite direction of the expected learning outcome of the Experimental Group as a result of attending two awareness sessions before taking the 3rd Test. This unexpected finding will be further discussed in the Chapter 6 Results Discussion. According to the significant difference being in the opposite direction of the experimental group, the effect size of the difference is also negative and medium sized. The effect size of this mean score difference between the two groups' is -0.35 according to Cohen's d (Trending Sideways, 2017). Cohen's d of -0.35 is considered negative medium size effect according to the range "> 0.2 and <= 0.5" (Cohen, 1992; Polyu.edu.hk, 2017).

Table 27 - Paired t-test comparing the scores of all participants combined between 2nd and 3rd test



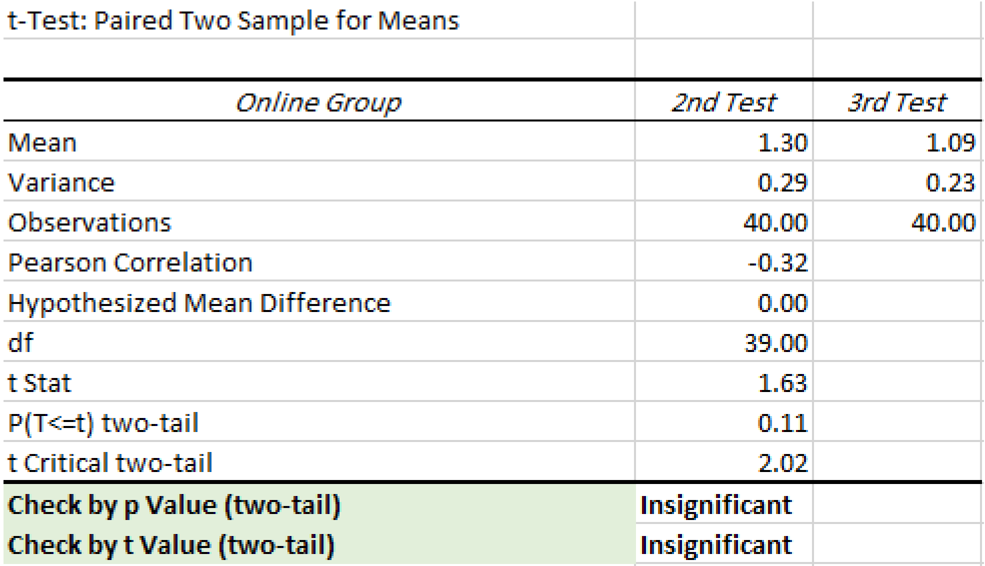
t-Test: Paired Two Sample for Means		
All Particpants	*2nd Test*	*3rd Test*
Mean	1.18	0.98
Variance	0.30	0.33
Observations	100.00	100.00
Pearson Correlation	-0.08	
Hypothesized Mean Difference	0.00	
df	99.00	
t Stat	2.39	
P(T<=t) two-tail	0.02	
t Critical two-tail	1.98	
Check by p Value (two-tail)	**Significant**	
Check by t Value (two-tail)	**Significant**	

H11: Comparing datasets 9 and 6

Null Hypothesis (11a): There is no statistically significant difference between the mean scores on the 3rd Test and the 2nd Test for the Online participant group

Test Results: A two tailed paired t-test was conducted to compare the mean scores between the 2nd Test and 3rd Test for the Online participant group to determine if there was a significant difference between the two tests and on which test, scores were significantly higher than the other. This is to assess whether the Online participant Group had benefited from the awareness session they attended after they had completed the 2nd Test and before they had taken the 3rd Test. The participants' scores on the 3rd Test (Experimental Group) are expected to be significantly higher than those on the 2nd Test (Control Group) after attending the awareness session. The result of the test shows as in Table 28 a non-significant difference between the means scores on the 2nd Test (M= 1.30, SD= 0.54) and the 3rd Test (M= 1.09, SD= 0.48) where the t statistic (df = 39) = 1.63 and the P value two-tail = 0.11 > alpha 0.05. Therefore, it is concluded that the difference between the mean scores on the 2nd and 3rd Tests for the Online participant group was not significantly different. Therefore, the null hypothesis (H11a) is supported and thus the alternate hypothesis (H11b) is rejected.

Table 28 - Paired t-test comparing the scores of Online Group between 2nd and 3rd test



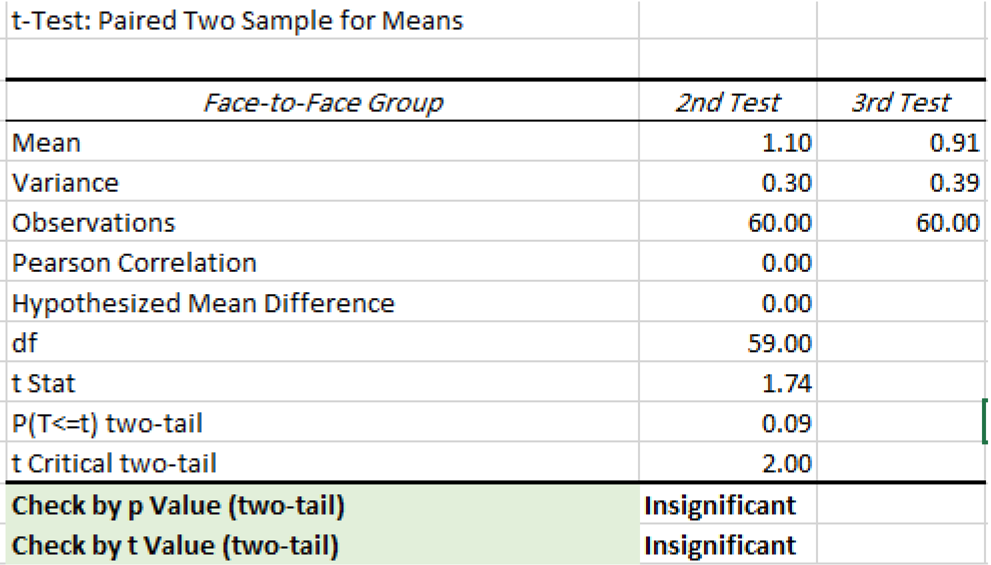
t-Test: Paired Two Sample for Means		
Online Group	*2nd Test*	*3rd Test*
Mean	1.30	1.09
Variance	0.29	0.23
Observations	40.00	40.00
Pearson Correlation	-0.32	
Hypothesized Mean Difference	0.00	
df	39.00	
t Stat	1.63	
P(T<=t) two-tail	0.11	
t Critical two-tail	2.02	
Check by p Value (two-tail)	**Insignificant**	
Check by t Value (two-tail)	**Insignificant**	

H12: Comparing datasets 8 and 5

Null Hypothesis (H12a): There is no statistically significant difference between the mean scores on the 3rd Test and the 2nd Test for the Face-to-Face participant group.

Test Results: A two tailed paired t-test was conducted to compare the mean scores between the 2nd Test and 3rd Test for the Face-to-Face participants to determine if there was a significant difference between the two tests and on which test, scores were significantly higher than the other. This is to assess whether the participants had benefited from the 2 awareness sessions they attended after they had completed the 2nd Test and before they had taken the 3rd Test. The participants' scores on the 3rd Test (Experimental Group) are expected to be significantly higher than those on the 2nd Test (Control Group) after attending the awareness sessions. The result of the test shows as in Table 29 a non-significant difference between the means scores on the 2nd Test (M= 1.10, SD= 0.54) and the 3rd Test (M= 0.91, SD= 0.62) where the t statistic (df = 56) = 1.74 and the P value two-tail = 0.09 > alpha 0.05. Therefore, it is concluded that the difference between the mean scores on the 2nd and 3rd Tests for the Face-to-Face participant group was not significantly different. This conclusion supports the Null hypothesis (H12a) and hence rejects the Alternate Hypothesis (H12b).

Table 29 - Paired t-test comparing the scores of Face-to-Face Group between 2nd and 3rd test



t-Test: Paired Two Sample for Means		
Face-to-Face Group	*2nd Test*	*3rd Test*
Mean	1.10	0.91
Variance	0.30	0.39
Observations	60.00	60.00
Pearson Correlation	0.00	
Hypothesized Mean Difference	0.00	
df	59.00	
t Stat	1.74	
P(T<=t) two-tail	0.09	
t Critical two-tail	2.00	
Check by p Value (two-tail)	**Insignificant**	
Check by t Value (two-tail)	**Insignificant**	

5.19 Score Variance between the 3rd and 1st Tests for each Participant Group

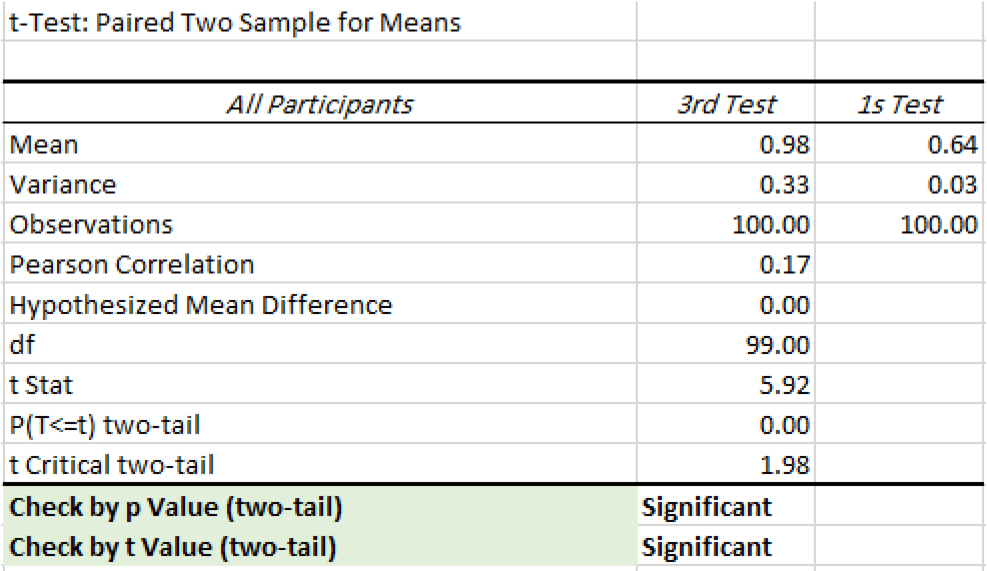
Finally, the next 3 hypotheses will assess the differences in mean scores between the 3rd and 1st Tests in the same pattern as the previous 6 hypotheses above assessing the mean scores between the 2nd and 3rd Tests and the 1st and 2nd Tests. Throughout the next 3 hypotheses, the effect of the 1st and 2nd Awareness sessions participants attended after the 1st and 2nd Tests and before taking the 3rd Test is assessed and measured.

H13: Comparing datasets 7 and 1

Null Hypothesis (H13a): There is no statistically significant difference between the mean scores on the 3rd Test and the 1st Test for all participants.

Test Results: A two tailed paired t-test was conducted to compare the mean scores between the 3rd Test and 1st Test for all participants to determine if there was a significant difference between the two tests and on which test, scores were significantly higher than the other. This is to assess whether the participants had benefited from the 2 awareness sessions they attended after they had completed the 1st and 2nd Tests and before they had taken the 3rd Test. The participants' scores on the 3rd Test (Experimental Group) are expected to be significantly higher than those on the 1st Test (Control Group) after attending the awareness sessions. The result of the test shows as in Table 30 a significant difference between the means scores on the 3rd Test (M= 0.98, SD= 0.57) and the 1st Test (M= 0.64, SD= 0.17) where the t statistic (df = 99) = 5.92 and the P value two-tail = (< 0.001) < alpha 0.05. Therefore, it is concluded that the difference between the mean scores on the 1st and 3rd Tests is significant where participants (Experimental Group) scored higher on the 3rd Test than on the 1st Test considering the higher mean score on the 3rd Test (3rd Test's M= 0.98 > 1st Test's M= 0.64) with a large positive effect size explained below. This conclusion does not support the Null hypothesis (H13a) and hence supports the Alternate Hypothesis (H13b). The effect size of the significance of this mean score difference between the two groups' is 0.90 according to Cohen's d (Trending Sideways, 2017). Cohen's d of 0.90 is considered a large size effect according to the range "> 0.8" (Cohen, 1992; Polyu.edu.hk, 2017).

Table 30 - Paired t-test comparing the scores of all participants combined between 1st and 3rd test



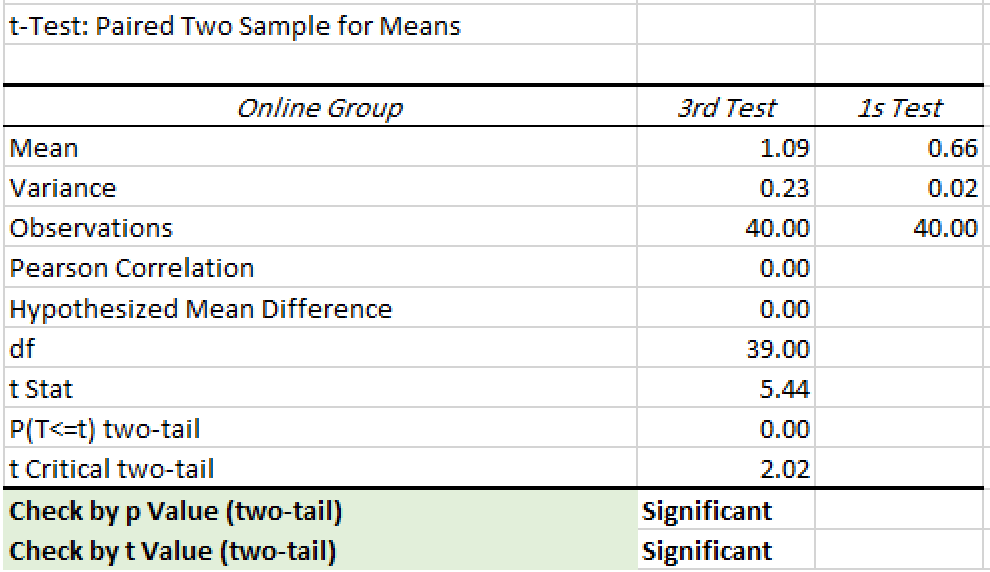
t-Test: Paired Two Sample for Means		
All Participants	*3rd Test*	*1s Test*
Mean	0.98	0.64
Variance	0.33	0.03
Observations	100.00	100.00
Pearson Correlation	0.17	
Hypothesized Mean Difference	0.00	
df	99.00	
t Stat	5.92	
P(T<=t) two-tail	0.00	
t Critical two-tail	1.98	
Check by p Value (two-tail)	**Significant**	
Check by t Value (two-tail)	**Significant**	

H14: Comparing datasets 9 and 6

Null Hypothesis (H14a): There is no statistically significant difference between the mean scores on the 3rd Test and the 1st Test for the Online participant group.

Test Results: A two tailed paired t-test was conducted to compare the mean scores between the 3rd Test and 1st Test for the Online participant group to determine if there was a significant difference between the two tests and on which test, scores were significantly higher than the other. This is to assess whether the participants had benefited from the 2 awareness sessions they attended after they had completed the 1st and 2nd Tests and before they had taken the 3rd Test. The participants' scores on the 3rd Test (Experimental Group) are expected to be significantly higher than those on the 1st Test (Control Group) after attending the awareness sessions. The result of the test shows as in Table 31 a significant difference between the means scores on the 3rd Test (M= 1.09, SD= 0.48) and the 1st Test (M= 0.66, SD= 0.14) where the t statistic (df = 39) = 5.44 and the P value two tail = (<0.001) < alpha 0.05. Therefore, it is concluded that the difference between the mean scores on the 1st and 3rd Tests is significant where participants (Experimental Group) scored higher on the 3rd Test than on the 1st Test (Control Group) considering the higher mean on the 3rd Test (3rd Test's M= 1.09 > 1st Test's M= 0.66) with a positive large effect size explained below. This conclusion does not support the Null hypothesis (H14a) and hence supports the Alternate Hypothesis (H14b). The effect size of the significance of this mean score difference between the two groups' is 1.38 according to Cohen's d (Trending Sideways, 2017). Cohen's d of 1.38 is considered a large size effect according to the range "> 0.8" (Cohen, 1992; Polyu.edu.hk, 2017).

Table 31 - Paired t-test comparing the scores of Online Group between 1st and 3rd test



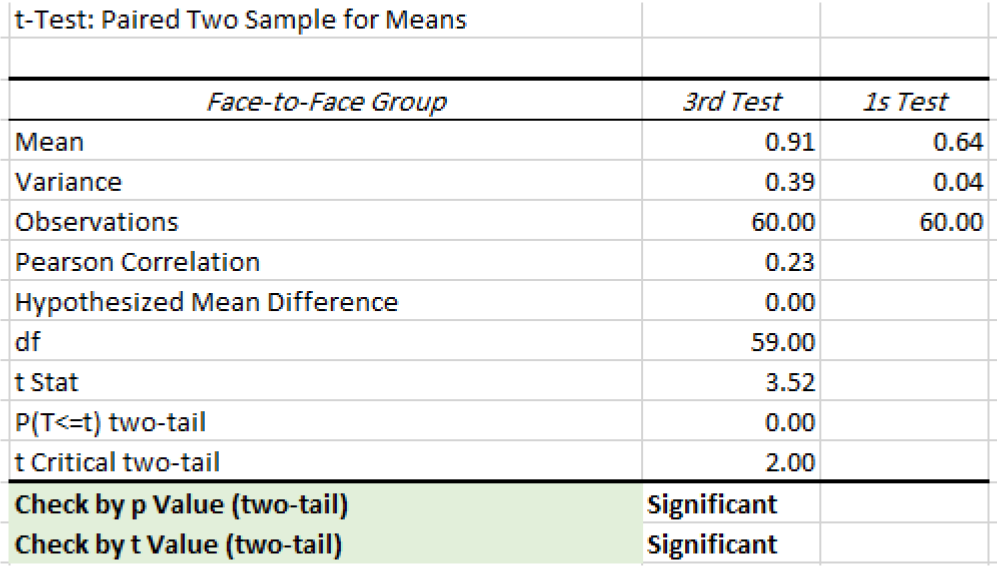
t-Test: Paired Two Sample for Means		
Online Group	*3rd Test*	*1s Test*
Mean	1.09	0.66
Variance	0.23	0.02
Observations	40.00	40.00
Pearson Correlation	0.00	
Hypothesized Mean Difference	0.00	
df	39.00	
t Stat	5.44	
P(T<=t) two-tail	0.00	
t Critical two-tail	2.02	
Check by p Value (two-tail)	**Significant**	
Check by t Value (two-tail)	**Significant**	

H15: Comparing datasets 8 and 5

Null Hypothesis (H15a): There is no statistically significant difference between the mean scores on the 3rd Test and the 1st Test for the Face-to-Face participant group.

Test Results: The same test is also applied to the Face-to-Face Group. Thus, a two tailed paired t-test was conducted to compare the mean scores between the 3rd Test and 1st Test for the Face-to-Face participant group to determine if there was a significant difference between the two tests and on which test, scores were significantly higher than the other. This is to assess whether the Face-to-Face participant group had benefited from the 2 awareness sessions they attended after they had completed the 1st and 2nd Tests and before they had taken the 3rd Test. The participants' scores on the 3rd Test (Experimental Group) are expected to be significantly higher than those on the 1st Test (Control Group) after attending the awareness sessions. The result of the test shows as in Table 32 a significant difference between the means scores on the 3rd Test (M= 0.91, SD= 0.62) and the 1st Test (M= 0.64, SD= 0.19) where the t statistic (df = 59) = 3.52 and the P value two-tail = (<0.001) < alpha 0.05. Therefore, it is concluded that the difference between the mean scores on the 1st and 3rd Tests is significant where participants (Experimental Group) scored higher on the 3rd Test than on the 1st Test (Control Group) considering the higher mean scores on the 3rd Test (3rd Test's M= 0.91 > 1st Test's M= 0.64) with a medium size effect explained below. This conclusion does not support the Null hypothesis (H15a) and hence supports the Alternate Hypothesis (H15b). The effect size of the significance of this mean score difference between the two groups' is 0.68 according to Cohen's d (Trending Sideways, 2017). Cohen's d of 0.68 is considered a medium size effect according to the range "> 0.5 and <0.8" (Cohen, 1992; Polyu.edu.hk, 2017).

Table 32 - Paired t-test comparing the scores of Face-to-Face Group between 1st and 3rd test



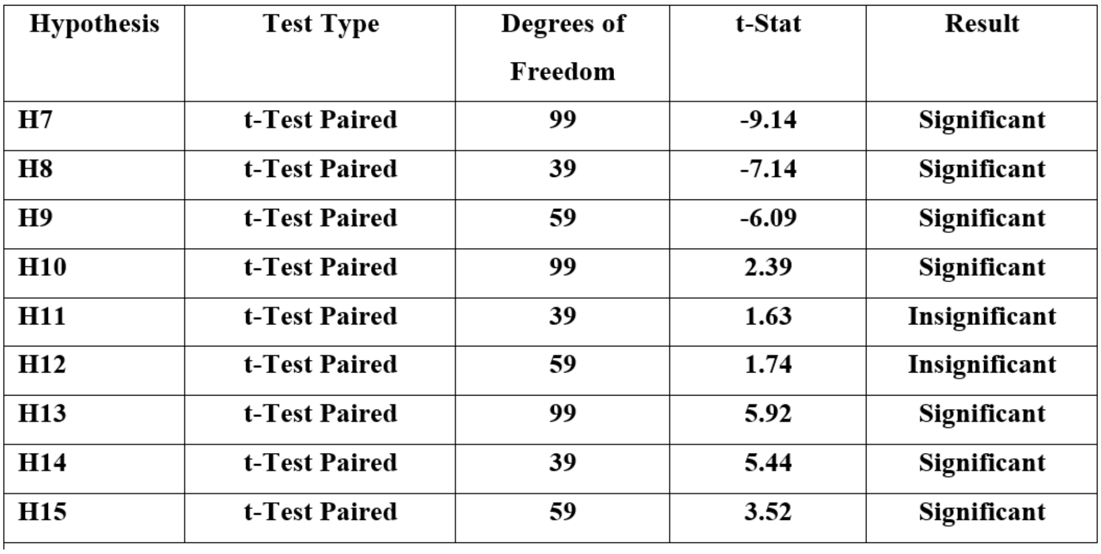
t-Test: Paired Two Sample for Means

Face-to-Face Group	*3rd Test*	*1s Test*
Mean	0.91	0.64
Variance	0.39	0.04
Observations	60.00	60.00
Pearson Correlation	0.23	
Hypothesized Mean Difference	0.00	
df	59.00	
t Stat	3.52	
P(T<=t) two-tail	0.00	
t Critical two-tail	2.00	
Check by p Value (two-tail)	**Significant**	
Check by t Value (two-tail)	**Significant**	

In summary as in Table 33, three hypotheses (H7, H8 and H9) compared 1st Test's scores with the 2nd Test' scores which indicated significant differences in the participants' mean scores and level of awareness being higher on the 2nd Test (Experimental Group) than that on the 1st Test (Control Group). The same results were found for the Online participant group, the Face-to-Face participants and all participants combined. Likewise three hypotheses (H13, H14 and H15) compared the 1st Test's scores with the 3rd Test's scores and indicated the same significant results where the experimental group outdid the control group on the 3rd Test for all participants, Online participant Group and Face-to-Face participant Group.

However, the three hypotheses (H10, H11 and H12) that compared the 2nd Test's scores with the 3rd Test's scores indicated significant differences in scores only for all participants combined but in the opposite direction where the Control Group on the 2nd Test scored higher than the Experimental Groups on the 3rd Test. The Experimental group whose learning outcome was expected to be higher and thus have higher scores on the 3rd Test than on the 2nd Test did not materialize according to this test. On the other hand, there were no significant differences in scores for the individual Online participant group and the Face-to-Face participant group.

Table 33- Results Summary for H7-H15



Hypothesis	Test Type	Degrees of Freedom	t-Stat	Result
H7	**t-Test Paired**	**99**	**-9.14**	**Significant**
H8	**t-Test Paired**	**39**	**-7.14**	**Significant**
H9	**t-Test Paired**	**59**	**-6.09**	**Significant**
H10	**t-Test Paired**	**99**	**2.39**	**Significant**
H11	**t-Test Paired**	**39**	**1.63**	**Insignificant**
H12	**t-Test Paired**	**59**	**1.74**	**Insignificant**
H13	**t-Test Paired**	**99**	**5.92**	**Significant**
H14	**t-Test Paired**	**39**	**5.44**	**Significant**
H15	**t-Test Paired**	**59**	**3.52**	**Significant**

5.20 Testing Independent Variables' Impact on Scores

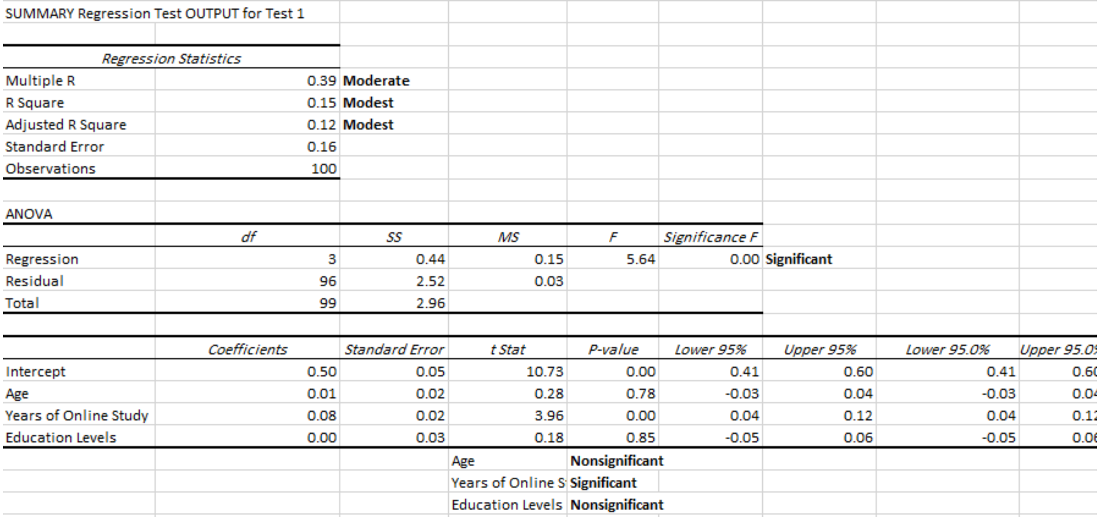
The last three hypotheses test the impact of a set of independent variables such as the participants' age groups, education levels and years of online study on the participants' level of awareness reflected by the scores on the 1st, 2nd and 3rd Tests respectively.

H16: Testing dataset 1

Null Hypothesis (H16a): The independent variables (Participants' Age, Education level and Years of online study) do not significantly impact the scores (dependent variable) on the 1st Test.

Test Results: A simple linear regression was calculated to predict participants' scores on the 1st Test based on three independent variables (1 participants' age group, 2) years of online study and 3) education level. As a result, a significant regression equation was found (F (3, 96) = 5.63, F value = 0.001 < alpha 0.05), with an R Square of 0.149 (Modest relationship: <= 0.3) which explains about 15% of the variations. The individual calculated P values for all the independent variables' coefficients (Age P Value = 0.78 > alpha 0.05 and Education Level P Value = 0.85 > alpha 0.05) were non-significant > 0.05 while the 'Years of Online Study' variable whose P value was (<0.001) < alpha = 0.05 shows significance. Therefore, participants' predicted scores will be predicted based on the following equation built from the Regression Test Results generated in Table 34: (Intercept) + (Coefficient of Age Group) * (Participant's Age Group) + (Coefficient of Years of Online Study) * (Participant's Years of Online Study) + (Coefficient of Education Levels) * (Participant's Education Level).

Table 34 - Simple Linear Regression and Anova tests to calculate participants' scores based on their 1st Test scores and associated coefficients



SUMMARY Regression Test OUTPUT for Test 1

Regression Statistics		
Multiple R	0.39	**Moderate**
R Square	0.15	**Modest**
Adjusted R Square	0.12	**Modest**
Standard Error	0.16	
Observations	100	

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*	
Regression	3	0.44	0.15	5.64	0.00	**Significant**
Residual	96	2.52	0.03			
Total	99	2.96				

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0*
Intercept	0.50	0.05	10.73	0.00	0.41	0.60	0.41	0.6
Age	0.01	0.02	0.28	0.78	-0.03	0.04	-0.03	0.0
Years of Online Study	0.08	0.02	3.96	0.00	0.04	0.12	0.04	0.1
Education Levels	0.00	0.03	0.18	0.85	-0.05	0.06	-0.05	0.0
			Age	**Nonsignificant**				
			Years of Online S	**Significant**				
			Education Levels	**Nonsignificant**				

Applying the formula to values from the regression test results above, the formula is: (0.50) + (0.005) * (Participant's Age Group) + (0.078) * (Participant's 'Years of Online Study') + (0.004) * (Participant's Education Level). This indicates that participants' scores on the 1st Test can be predicted to increase by (0.078) which is about 8% (modest) for every higher category of the 'Years of online study' variable for each participant. However, the P values of the other two independent variables' coefficients (Age Coefficient = 0.005, P Value = 0.78 > alpha 0.05 and Education Level Coefficient = 0.004, P Value = 0.85 > alpha 0.05) are not considered significant predictors into the participants' scores on the 1st Test. Therefore, it is concluded from the results above that the Null hypothesis (H16a) is not supported since the 'Years of Online Study' variable has a modest significant impact on predicting participants' scores on the 1st Test and thus the Alternate Hypothesis (H16b) is supported. Residual and Line Fit plot graphs of the Years of Online Study, Education Levels and Age Groups are shown in Figure 40.

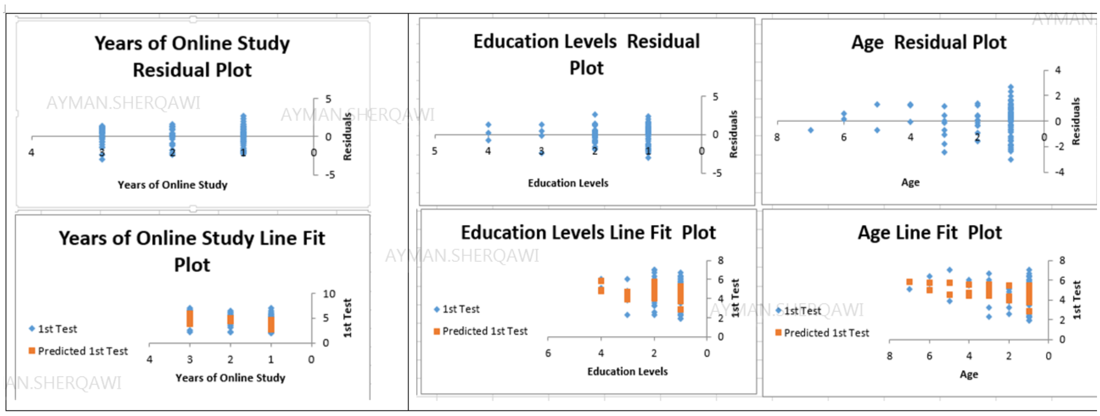


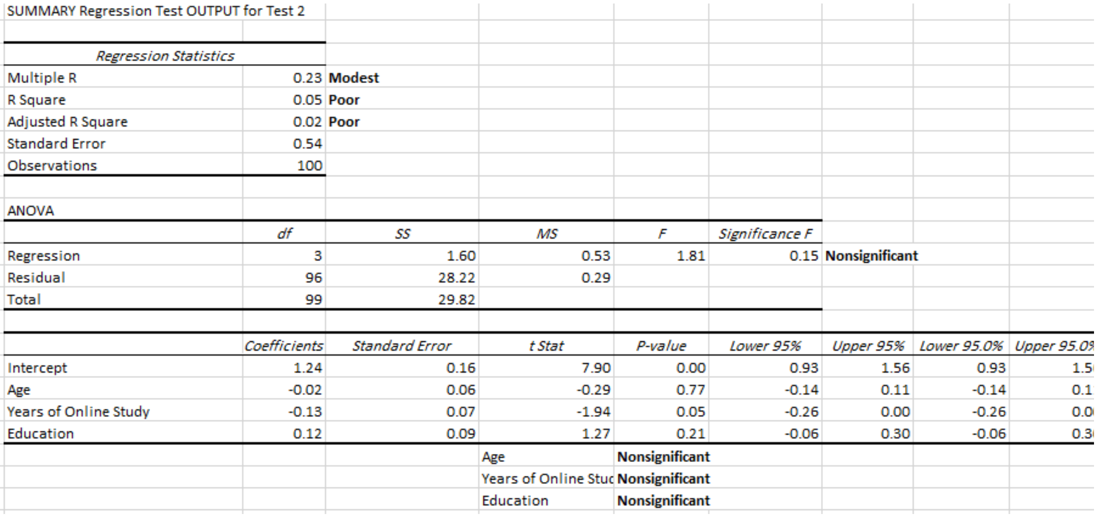
Figure 40 - Residual and Line Fit Plots for Participants' Years of Online Study, Education and Age variables

H17: Testing dataset 4

Null Hypothesis (H17a): The independent variables (Participants' Age, Education level and Years of online study) do not significantly impact the scores (dependent variable) on the 2nd Test.

Test Results: A simple linear regression was calculated to predict participants' scores on the 2nd Test based on three independent variables (1 participants' age group, 2) years of online study and 3) education level. The resulting ANOVA F Significance level was > 0.05 which indicated insignificance (F (3, 96) = 1.80, F value = 0.15 > alpha 0.05), with an R Square of 0.05 (Poor relationship: <= 0.1) which explains about 5% of the variations. In addition, the individual calculated P values for all the independent variables' coefficients (Age P Value = 0.77 > alpha 0.05, Years of Online Study P Value = 0.054 > alpha 0.05 and Education Level P Value = 0.20 > alpha 0.05) were all non-significant > 0.05 as shown in the Regression Test Results generated in Table 35.

Table 35 - Simple Linear Regression and Anova tests to calculate participants' scores based on their 2nd Test scores and associated coefficients



SUMMARY Regression Test OUTPUT for Test 2

Regression Statistics		
Multiple R	0.23	**Modest**
R Square	0.05	**Poor**
Adjusted R Square	0.02	**Poor**
Standard Error	0.54	
Observations	100	

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*	
Regression	3	1.60	0.53	1.81	0.15	**Nonsignificant**
Residual	96	28.22	0.29			
Total	99	29.82				

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0*
Intercept	1.24	0.16	7.90	0.00	0.93	1.56	0.93	1.5
Age	-0.02	0.06	-0.29	0.77	-0.14	0.11	-0.14	0.1
Years of Online Study	-0.13	0.07	-1.94	0.05	-0.26	0.00	-0.26	0.0
Education	0.12	0.09	1.27	0.21	-0.06	0.30	-0.06	0.3

Age	**Nonsignificant**
Years of Online Stuc	**Nonsignificant**
Education	**Nonsignificant**

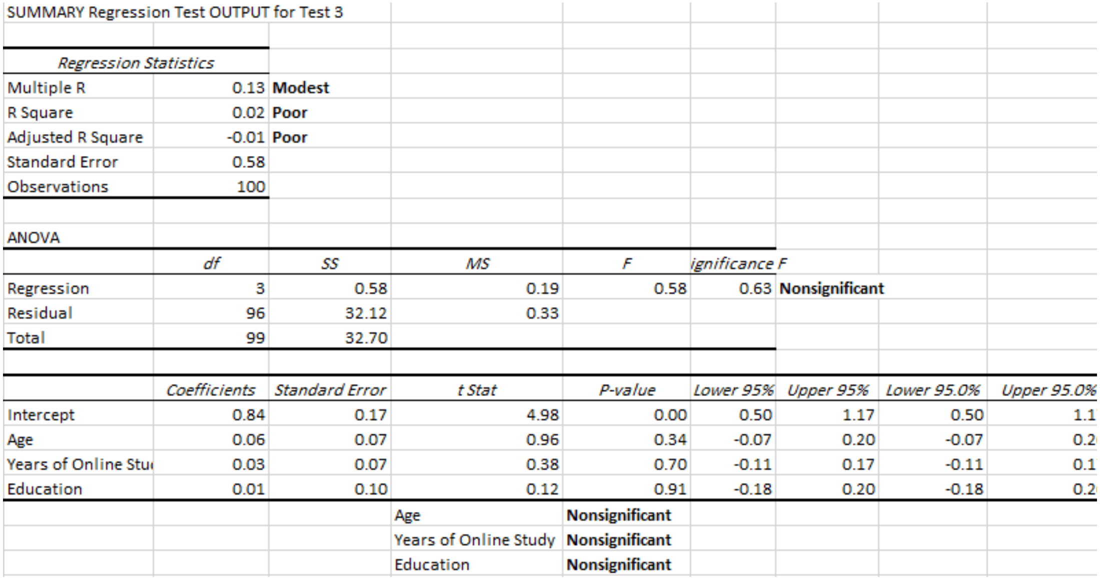
Therefore, it is concluded from the results above that the Null hypothesis (H17a) is supported since all the independent variables have a non-significant and poor predicting power on participants' scores on the 2nd Test and hence the Alternate Hypothesis (H17b) is rejected.

H18: Testing dataset 7

Null Hypothesis (H18a): The independent variables (Participants' Age, Education level and Years of online study) do not significantly impact the scores (dependent variable) on the 3rd Test.

Test Results: Finally, a simple linear regression was calculated to predict participants' scores on the 3rd Test based on three independent variables as in the above hypothesis. The resulting ANOVA F Significance level was > 0.05 which indicated insignificance (F (3, 96) = 0.58, F value = 0.62 > alpha 0.05), with an R Square of 0.01 (Poor relationship: <= 0.1) which explains about 1% of the variations. In addition, the individual calculated P values for all the independent variables' coefficients (Age P Value = 0.34 > alpha 0.05, Years of Online Study P Value = 0.70 > alpha 0.05 and Education Level P Value = 0.90 > alpha 0.05) were all non-significant > 0.05 as shown in the Regression Test Results generated in Table 36.

Table 36 - Simple Linear Regression and Anova tests to calculate participants' scores based on their 3rd Test scores and associated coefficients



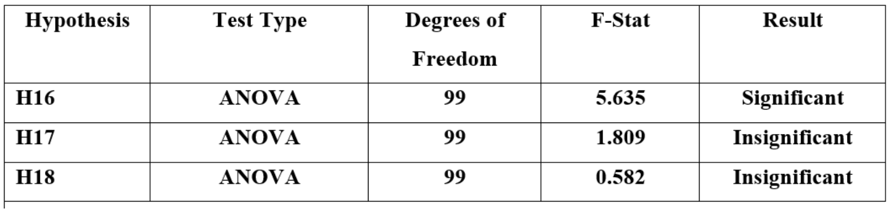
SUMMARY Regression Test OUTPUT for Test 3								
Regression Statistics								
Multiple R	0.13	**Modest**						
R Square	0.02	**Poor**						
Adjusted R Square	-0.01	**Poor**						
Standard Error	0.58							
Observations	100							
ANOVA								
	df	*SS*	*MS*	*F*	*ignificance F*			
Regression	3	0.58	0.19	0.58	0.63	**Nonsignificant**		
Residual	96	32.12	0.33					
Total	99	32.70						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	0.84	0.17	4.98	0.00	0.50	1.17	0.50	1.1
Age	0.06	0.07	0.96	0.34	-0.07	0.20	-0.07	0.2
Years of Online Stu	0.03	0.07	0.38	0.70	-0.11	0.17	-0.11	0.1
Education	0.01	0.10	0.12	0.91	-0.18	0.20	-0.18	0.2
			Age	**Nonsignificant**				
			Years of Online Study	**Nonsignificant**				
			Education	**Nonsignificant**				

Therefore, the Null hypothesis (18a) is supported since all the independent variables are non-significant with poor predicting power on participants' scores on the 3rd Test and hence the Alternate Hypothesis (18b) is rejected.

In summary of all the above, we conclude that H1, H2 and H3 were supported indicating no significant variances in the scores between the Online participant group and the Face-to-Face participant group on the 1st, 2nd and 3rd Tests respectively. On the other hand, H4, H5 and H6 were not supported indicating a significant difference between the scores of the 1st, 2nd and 3rd Tests for all participants combined and separated as the Online and Face-to-Face participant groups. Likewise, H7, H8 and H9 were not supported indicating significant differences between the 2nd Test and 1st Test scores for all participant groups. On the other hand, H10 was not supported, but indicated significance in the opposite direction of what was expected as a significantly higher mean score on the 3rd Test than on the 2nd Test for all participants combined, while H11 and H12 were supported indicating no significance for the individual Online and Face-to-Face participant groups. Nonetheless, H13, H14 and H15 were not supported indicating significant differences between the 3rd Test and 1st Test scores for all participants combined and separated as the Online and Face-to-Face participant groups. Similarly as shown in Table 37 below, H16 was not supported indicating that the higher number of years of online study participants had, the higher scores they are predicted to acquire on the 1st Test. However, the scores on the 2nd and 3rd Tests were not significantly impacted by the combination of participants' independent variables such as age, education level and years of online study as indicated by H17 and H18.

Although H17 and H18 did not provide much significance that allows participants' scores to be predicted based on these independent variables except for the Years of Online Study, Spearman's Rank-Order Correlation tests (Ramsey, 1989) were conducted to test for correlations between the increase and decrease of the scores in relation to 5 participants' demographical traits individually such as participants' age, years of online study, education level, elapsed days between attended awareness sessions and ISS role evaluation ratings. These tests and their implications with the other hypotheses' findings are discussed in detail next.

Table 37 - Results Summary of H16, H17 and H18

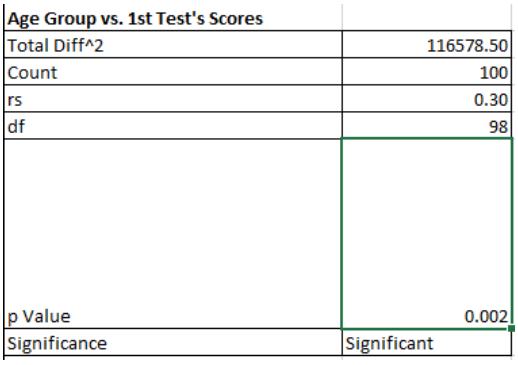


Hypothesis	Test Type	Degrees of Freedom	F-Stat	Result
H16	**ANOVA**	**99**	**5.635**	**Significant**
H17	**ANOVA**	**99**	**1.809**	**Insignificant**
H18	**ANOVA**	**99**	**0.582**	**Insignificant**

5.21 Correlations

1st Test′ Scores Correlation with Participant Age: A Spearman's rank-order correlation test was run to determine the relationship between 100 participants' scores on the 1st Test and their age. The test results show a strong, positive correlation between the scores on the 1st Test and the age of participants, which was statistically significant (Spearman Rank Correlation Coefficient "rs" = 0.300, p = 0.002 < α: 0.05) as shown in Table 38 and Figure 22. Therefore, it is concluded that the older participants were, the higher scores they acquired on the 1st Test.

Table 38 - Spearman's Rank-order Correlation test to determine relationship between 100 participants' scores on the 1st Test and their age groups



Age Group vs. 1st Test's Scores	
Total Diff^2	116578.50
Count	100
rs	0.30
df	98
p Value	0.002
Significance	Significant

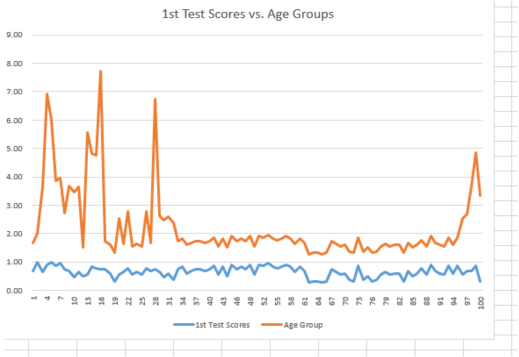
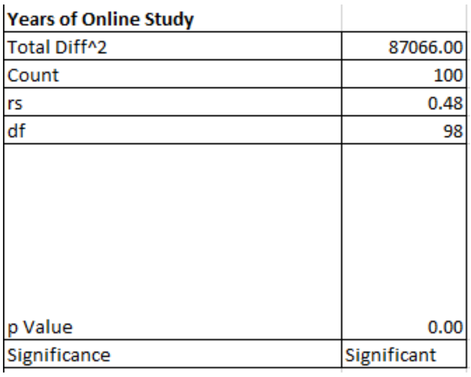


Figure 22 - relationship between 100 participants' scores on the 1st Test and their age groups

1st Test′ Scores Correlation with Number of Years in Online Study: As for the 'Years of Online Study' variable, the Spearman's rank-order correlation test gave a strong, positive correlation between the 1st Test's scores and the number of years of online study, which was statistically significant (Spearman Rank Correlation Coefficient "rs" = 0.477, p = (<0.001) < α: 0.05) as shown in Table 39 and Figure 20. Based on this result, it is concluded that the higher the number of years participants spent doing online study, the higher the scores they received on the 1st Test. These results agree with the results from the Regression test conducted earlier for the Alternate Hypothesis (H16b) on the 1st Tests' scores and their predictability based on the Years of Online Study independent variable.

Table 39 - Spearman's Rank-order Correlation test to determine relationship between 100 participants' scores on the 1st Test and their Years of Online Study variable



Years of Online Study	
Total Diff^2	87066.00
Count	100
rs	0.48
df	98
p Value	0.00
Significance	Significant

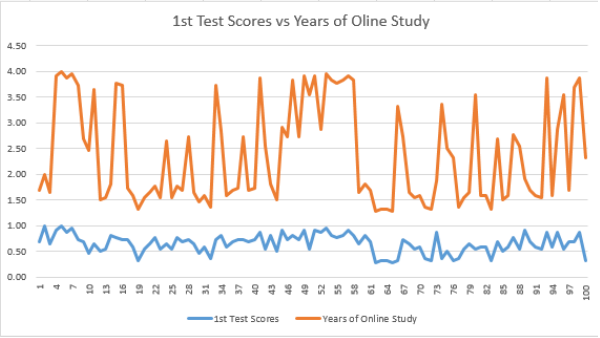
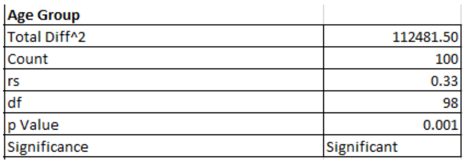


Figure 20 - Spearman's Rank-order Correlation test to determine relationship between 100 participants' scores on the 1st Test and their Years of Online Study variable

1st Test' Scores Correlation with Education Level and ISS Role Evaluation: On the other hand, the Spearman's Rank-Order correlation test was run against the participants' education levels (rs = 0.188, p = 0.06 > α: 0.05) and their ISS role evaluation ratings (rs = -0.143, p = 0.15 > α: 0.05). The tests' results were non-significant which are in line with the results of the previously conducted Regression tests on the education level variable and its impact on the scores of the 1st Test.

2nd Test' Scores Correlation with Participant Age: The Age Group variable is again shown to have a correlation with the participants' scores, but this time with the 2nd Test's scores. Thus, a Spearman's rank-order correlation was run to determine the relationship between 100 participants' scores on the 2nd Test and their age. The test shows a strong, positive correlation between the scores and the participant age, which was statistically significant (Spearman Rank Correlation Coefficient "rs" = 0.325, p = 0.0009 < α: 0.05) as shown in Table 40. Therefore, it is concluded that the older the participants were, the higher the scores they acquired on the 2nd Test as shown in Table 40 and Figure 21.

Table 40 -Spearman's Rank-order Correlation test to determine relationship between 100 participants' scores on the 2nd Test and their Age groups



Age Group	
Total Diff^2	112481.50
Count	100
rs	0.33
df	98
p Value	0.001
Significance	Significant

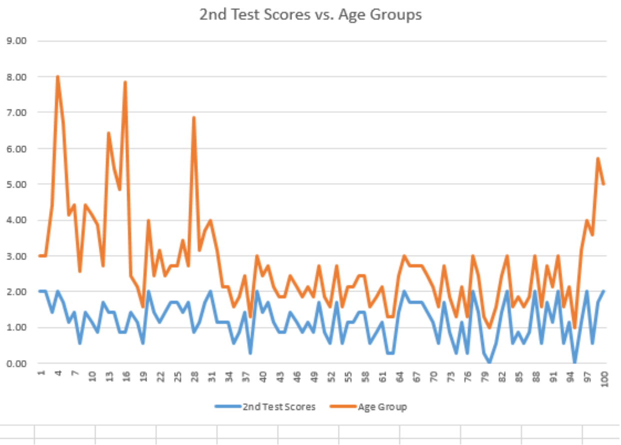
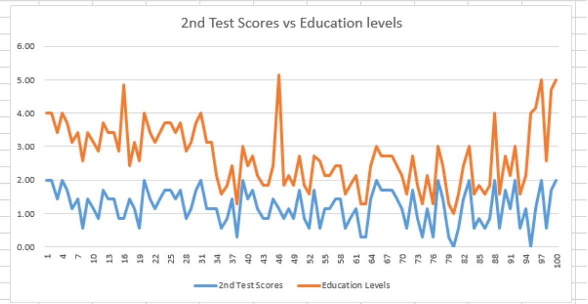


Figure 21 - Spearman's Rank-order Correlation test to determine relationship between 100 participants' scores on the 2nd Test and their Age groups

2nd Test' Scores Correlation with Number of Years in Online Study and Education Levels: On the other hand, the Years of Online Study does not have a significant correlation with the 2nd Test's scores since the resulting p-Value = (0.66 > 0.05). Unlike the 1st Test, the participants' education levels have a positive correlation with the 2nd Test's scores which was statistically significant (Spearman Rank Correlation Coefficient "rs" = 0.322, p = 0.001 < α: 0.05) as shown in Table 41. As a result, it is concluded that the higher the participant education levels were, the higher the scores they acquired on the 2nd Test as shown in Table 41 and Figure 22.

Table 41 - Spearman's Rank-order Correlation test to determine relationship between 100 participants' scores on the 2nd Test and their Education levels

Education Levels	
Total Diff^2	112851.00
Count	100
rs	0.32
df	98
p Value	0.001
Significance	Significant



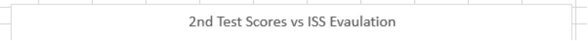
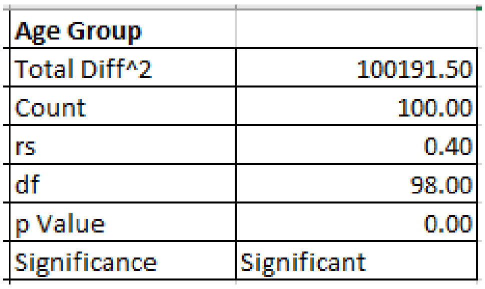


Figure 22 - Spearman's Rank-order Correlation test to determine relationship between 100 participants' scores on the 2nd Test and their Education levels

2nd Test' Scores Correlation with Awareness Session Interval and ISS Role Evaluation: On the other hand, the ISS Role evaluation ratings (rs = -0.118, p = 0.238 > α: 0.05) and the Elapsed number of days between attended awareness sessions (rs = -0.038, p = 0.704 > α: 0.05) do not have any significant correlations with the 2nd Test's scores.

3rd Test' Scores Correlation with Participant Age: Finally, Correlational relationships between the independent variables and the 3rd Test's scores are tested and analysed. Again and for the third time, participant age has a significant correlation with the scores on the 3rd Test which was statistically significant (Spearman Rank Correlation Coefficient "rs" = 0.398, p = (<0.001) < α: 0.05) as shown in Table 42. Therefore, it is concluded that the older participants were, the higher the scores they acquired on the 3rd Test as shown in Table 42 and Figure 23.

Table 42 - Spearman's Rank-order Correlation test to determine relationship between 100 participants' scores on the 3rd Test and their Age groups



Age Group	
Total Diff^2	100191.50
Count	100.00
rs	0.40
df	98.00
p Value	0.00
Significance	Significant

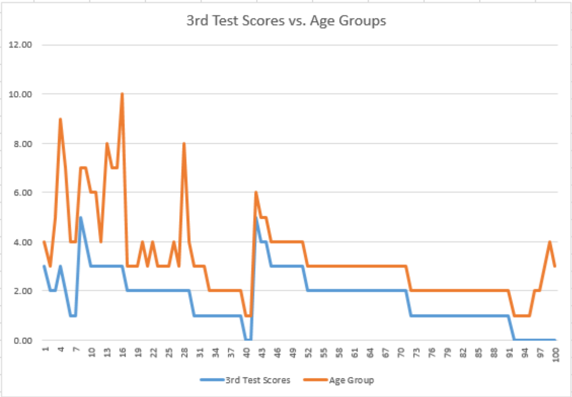
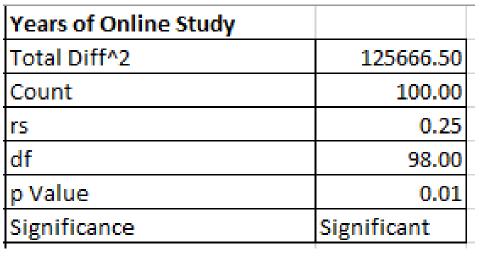


Figure 23 - Spearman's Rank-order Correlation test to determine relationship between 100 participants' scores on the 3rd Test and their Age groups

3rd Test' Scores Correlation with Number of Years in Online Study: Running the Spearman correlational test on the Years of Online Study variable, results in a strong, positive correlation between the 3rd Test's scores and the number of online study years, which was statistically significant (Spearman Rank Correlation Coefficient "rs" = 0.245, p = 0.013 < α: 0.05) as shown in Table 43. It is concluded that the higher the number of years participants spent doing online study, the higher the scores they received on the 3rd Test as shown in Table 43 and Figure 24.

Table 43 - Spearman's Rank-order Correlation test to determine relationship between 100 participants' scores on the 3rd Test and their Years of Online Study



Years of Online Study	
Total Diff^2	125666.50
Count	100.00
rs	0.25
df	98.00
p Value	0.01
Significance	Significant

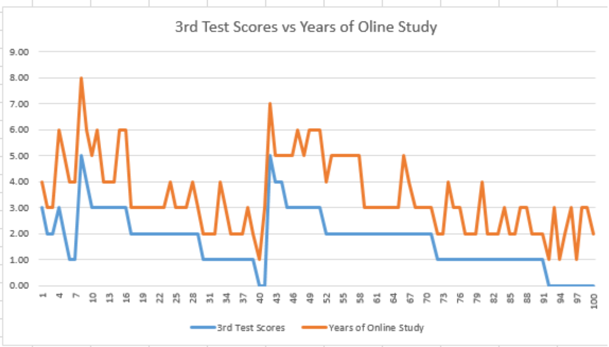
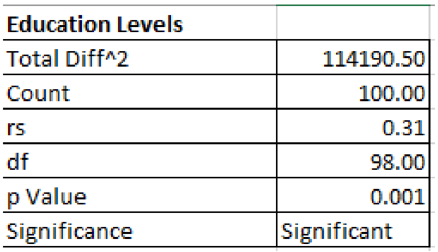


Figure 24 - Spearman's Rank-order Correlation test to determine relationship between 100 participants' scores on the 3rd Test and their Years of Online Study

3^{rd} Test' Scores Correlation with Participant's Education Level: As shown in Table 44 and Figure 25, a Spearman's rank-order correlation test was run to determine the relationship between 100 participants' scores on the 3^{rd} Test and their education levels. The test's results show a strong, positive correlation between the scores and participants' education levels, which was statistically significant (Spearman Rank Correlation Coefficient "rs" = 0.314, p = 0.001 < α: 0.05). Therefore, it is concluded that the higher the education levels participants had, the higher the scores they acquired on the 3^{rd} Test.

Table 44 - Spearman's Rank-order Correlation test to determine relationship between 100 participants' scores on the 3rd Test and their Education levels



Education Levels	
Total Diff^2	114190.50
Count	100.00
rs	0.31
df	98.00
p Value	0.001
Significance	Significant

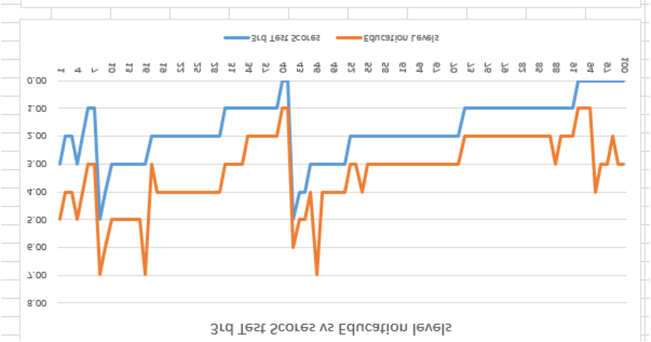
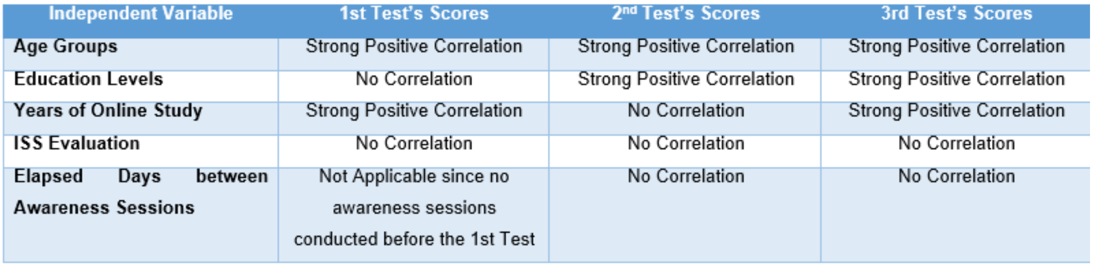


Figure 25 - Spearman's Rank-order Correlation test to determine relationship between 100 participants' scores on the 3rd Test and their Education levels

3rd Test' Scores Correlation with Awareness Session Interval and ISS Role Evaluation: The ISS Role Evaluation rating variable (rs = -0.105, p = 0.296 > α: 0.05) and the Elapsed days between attended awareness sessions (rs = -0.035, p = 0.722 > α: 0.05) do not have any significant correlations with the scores on the 3rd Tests. In summary, Table 45 summarizes the tests' results for the five independent variables and their correlations to the scores of the 3 Tests respectively.

Table 45 - Summary of the correlation tests' results for the five independent variables on the scores of the 3 Tests



Independent Variable	1st Test's Scores	2nd Test's Scores	3rd Test's Scores
Age Groups	Strong Positive Correlation	Strong Positive Correlation	Strong Positive Correlation
Education Levels	No Correlation	Strong Positive Correlation	Strong Positive Correlation
Years of Online Study	Strong Positive Correlation	No Correlation	Strong Positive Correlation
ISS Evaluation	No Correlation	No Correlation	No Correlation
Elapsed Days between Awareness Sessions	Not Applicable since no awareness sessions conducted before the 1st Test	No Correlation	No Correlation

The Age Groups, Education Levels and Years of Online Study variables have correlational relationships with the scores on almost all the three tests, while the ISS Evaluation and the Elapsed days between attended awareness sessions do not have any significant correlations with the scores on any of the three tests. As shown in the summary above, the Age Groups variable has a strong positive correlation with the scores on all the three tests. The education levels show the same correlational relationship as that of the age groups, but only with the scores on the 2nd and 3rd Tests. In the same sense, the Years of Online study variable shows a strong positive correlation with the scores on the 1st and 3rd Tests only.

5.22 Awareness Risk Residual Categories

The 18 hypotheses tested have provided answers as to whether the proposed SAMFP model has achieved its goals by improving the participants' level of awareness. However, the analysis does not cater for measuring the gap or the awareness risk residual between the participants' achieved awareness levels and the highest awareness level achievable which is Projection. Therefore, the Information Security Awareness and Capability Model (ISACM) of Poepjes (2012) is capitalized upon to measure this gap. The ISACM provides through one of its three attributes, the Awareness Risk; a mechanism to measure this awareness risk residual by calculating the difference between the Risk Capability which is the achieved awareness level represented by the participant's final score on all the three Tests (Sum of the three Tests' scores) and the Awareness Importance which is the weight assigned to the highest level of awareness targeted. Thus, applying this mechanism by summing the weighted scores on all the three tests for each participant and subtracting the total (final score) from the maximum weight assigned to the highest level of awareness targeted (Projection level), which has a weight of 7 out of 7. Then, the resulting differences are categorized according to their equality or proximity to any of the three weights (1, 3 and 7) respective to the three awareness levels (Perception, Comprehension and Projection). In other words, if the final score of a participant is less than or equal to 1 which is the weight assigned to the Perception Level, the respective Awareness Risk is categorized as a 'Perception Level Risk', whereas if the final score is greater than 1 and equal to or less than 3 which is the weight assigned to the Comprehension level, the respective Awareness Risk is categorized as a 'Comprehension Level Risk' and finally, if the final score is greater than 3 and equal to or less than 7 which is the weight assigned to the Projection level, the respective Awareness Risk is categorized as a 'Projection Level Risk'.

These Awareness Risk categorizations mean that the participants whose Awareness Risk is within a lower category might be exposed to Phishing Risks mitigatable by higher awareness levels. For example a participant whose Awareness Risk is categorized as ‘Comprehension Level Risk’ may be protected from the ‘Perception Level Risk which is a lower risk category, but is prone to the ‘Projection Level Risk’ which is the higher risk category than his/her current achieved awareness level. Figures 26 and 27 illustrate the percentages of participants’ distribution over the three awareness levels’ risk categories.

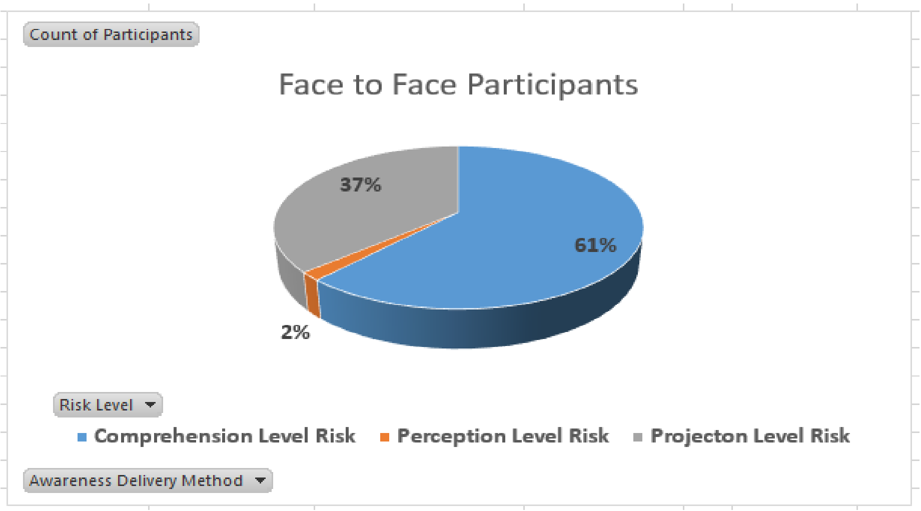


Figure 26 - Calculated Situation Awarness Levels' Risk for the Face-to-Face participants

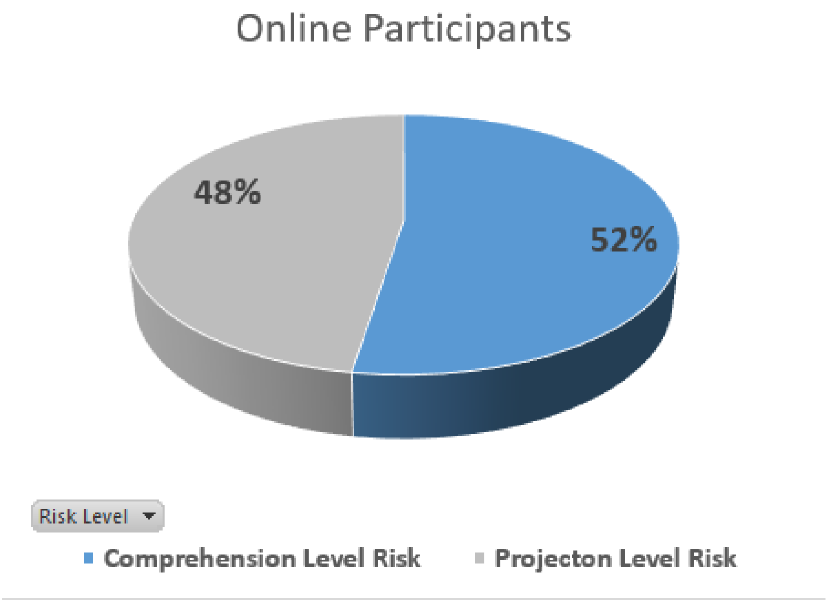


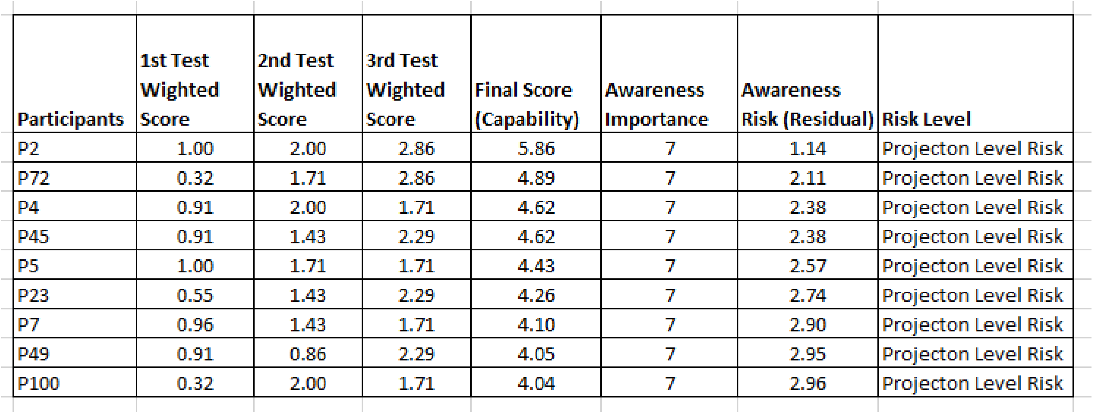
Figure 27 - Calculated Situation Awareness Levels' Risk for the Online participants

As explained above, the biggest percentage of the 100 participants (58%) have achieved the Comprehension Level according to the SAMFP Model and hence have the Comprehension Level Risk category which is a middle Risk. On the other hand, 41% of the participants have achieved the highest awareness level on SAMFP Model which is the Projection level risk and therefore have the lowest Awareness Risk level. Only 1% of the participants have the highest Awareness Risk Level relative to the Perception level on the SAMFP Model. These results indicate that more than half of the participants have made it to the second level of awareness. Yet, they are exposed to the Projection Level Risk and therefore, need to have their awareness enhanced to achieve the highest level of awareness possible which is Projection. Looking at participants' awareness risk levels based on their participant groups, Figures 28 and 29 show that 61% of the Face-to-Face participants have achieved the 'Comprehension Level' and therefore are at the Comprehension Level Risk as opposed to only 52% of the Online participants are at the same Risk Level. On the other hand, 37% of the Face-to-Face participants as opposed to 48% of the Online participants are exposed to the Projection Level Risk which is the lowest risk level. Finally, only 1% of the Face-to-Face participants is exposed to the Perception Level Risk which is the highest level of awareness risk as opposed to none from the Online participants. These percentages indicate that more of the Online participants than those of the Face-to-Face participants have achieved better levels of awareness and therefore are exposed to lower risk levels.

5.23 Understanding the Attributes of Highest Ranked Participants

Now the highest ranked participants who achieved the highest level of Phishing awareness and have the lowest risk categories are those who scored between 4 and 7 out of 7 as far as their final score after summing their scores from all the three tests is concerned. This criteria resulted in 9 participants selected as the highest ranked 9 participants with the highest scores and lowest risk levels. The highest score recorded was 5.86 out of 7 with the lowest risk level as 1.14 whereas the lowest score recorded in the highest score range (4-7) was 4.04 with a risk level of 2.96. All the highest ranked 9 participants have achieved the Projection awareness level and therefore have a Projection Risk Level which ranges from 0 to 3. The first highest ranked participant has accordingly the lowest risk level in the Projection Level Risk range ever recorded which was 1.14, while the 9th (last) participant has the highest Projection Risk Level ever recorded in the range which was 2.96. Table 46 lists the highest ranked 9 participants with their weighted scores and risk levels.

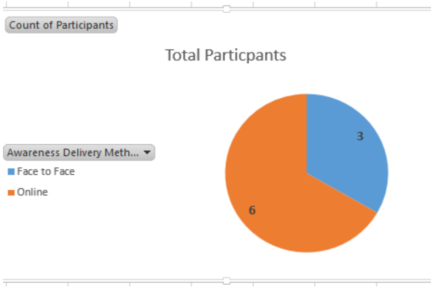
Table 46 - Highest ranked participants List



Participants	1st Test Wighted Score	2nd Test Wighted Score	3rd Test Wighted Score	Final Score (Capability)	Awareness Importance	Awareness Risk (Residual)	Risk Level
P2	1.00	2.00	2.86	5.86	7	1.14	Projecton Level Risk
P72	0.32	1.71	2.86	4.89	7	2.11	Projecton Level Risk
P4	0.91	2.00	1.71	4.62	7	2.38	Projecton Level Risk
P45	0.91	1.43	2.29	4.62	7	2.38	Projecton Level Risk
P5	1.00	1.71	1.71	4.43	7	2.57	Projecton Level Risk
P23	0.55	1.43	2.29	4.26	7	2.74	Projecton Level Risk
P7	0.96	1.43	1.71	4.10	7	2.90	Projecton Level Risk
P49	0.91	0.86	2.29	4.05	7	2.95	Projecton Level Risk
P100	0.32	2.00	1.71	4.04	7	2.96	Projecton Level Risk

It is observed that 6 participants out of the 9 (P2, P72, P45, P5, P23 and P7) have made steady improvements in their (weighted) scores from the 1st Test through the 3rd Test. This indicates that the awareness improvements they had achieved were gradual and in line with the proposed SAMFP awareness model. This means that participants learn gradually from the awareness sessions they had attended as indicated by their scores on the three assessment tests. Accordingly, these assessment tests were designed with gradual levels of difficulty to test if participants made meaningful progress in their awareness levels about Phishing e.g. participants improved from Perception level to Comprehension Level and from Comprehension Level to the Projection Level.

Probing deeper to understand if there was something special about these 9 participants, it has been found that 6 participants out of the 9 belong to the Online participant group, while the other 3 belong to the Face-to-Face participant group as shown in Figures 28 and 29.



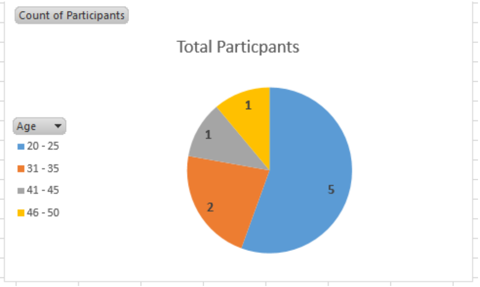


Figure 28 - Highest ranked participants by participant groups

Figure 29 - Highest ranked participants by Age Group

From an age perspective, 5 participants out of the 9 are between 20 to 25 years old, 2 of them are between 31 to 35 years old, while the other 2 participants are in the 41-45 and 46-50 age groups respectively as illustrated in Figure 31. This indicates that most of the highest ranked participants are at a young age.

From the education perspective, again 5 participants are studying for a bachelor degree, while only 3 are at the high school level or below with the last participant being at the Master's degree level. This indicates that the majority of the highest ranked 9 participants are at the Bachelor Degree level as shown in Figure 30.

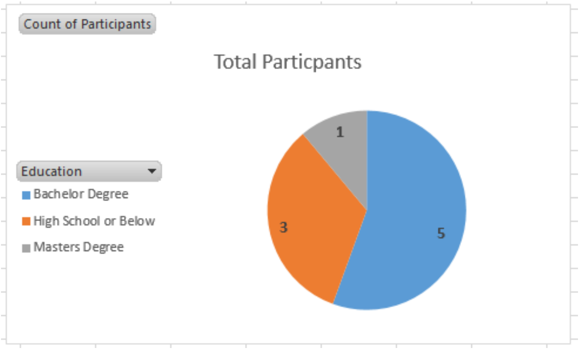


Figure 30 - Highest ranked participants by Education levels

Although, almost half of the 9 participants (4 participants) have spent 5 years or more in online study, the first highest ranked participant has only been an online student for less than one year along with 2 others making a total of 3 out of the 9 participants. Finally, only 2 participants have spent 3 years in online study as n Figure 31. This does not indicate a consistent relationship between the number of years of online study and the 9 participants' highest scores. Only this small subset of the data does not go in line with the results of the Regression tests conducted previously on all participants which indicated correlational relationships between participants' number of years in online study and the scores on the 1st Test as an example.

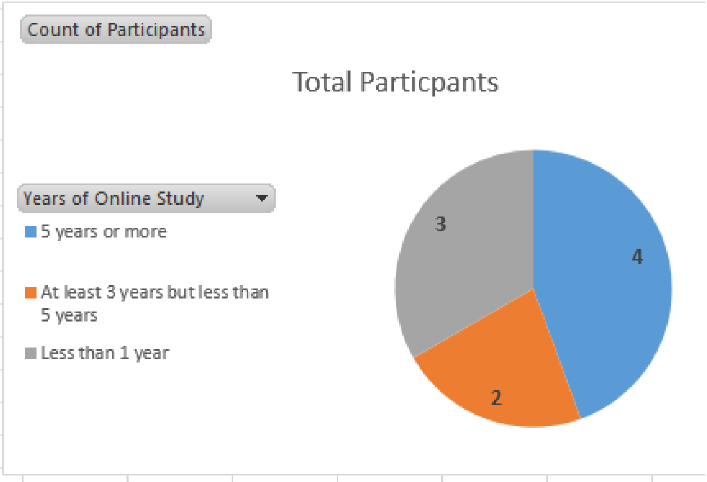


Figure 31 - Highest ranked participants by Number of Years of Online Study

As Figure 32 shows, there are 4 participants out of the 9 participants have evaluated the role of their ISS as the best on a scale of 0 to 5. Contrary to this is that the highest ranked participant has evaluated the role of ISS as the 'Worst'. The remaining 4 participants have rated their ISS between 2, 3 and 4. Again, Figure 34 does not indicate a pattern of any correlation between the highest ranked participant's evaluation rates of their ISS and their highest scores. This goes in line with the results of the Regression tests conducted previously on all participants.

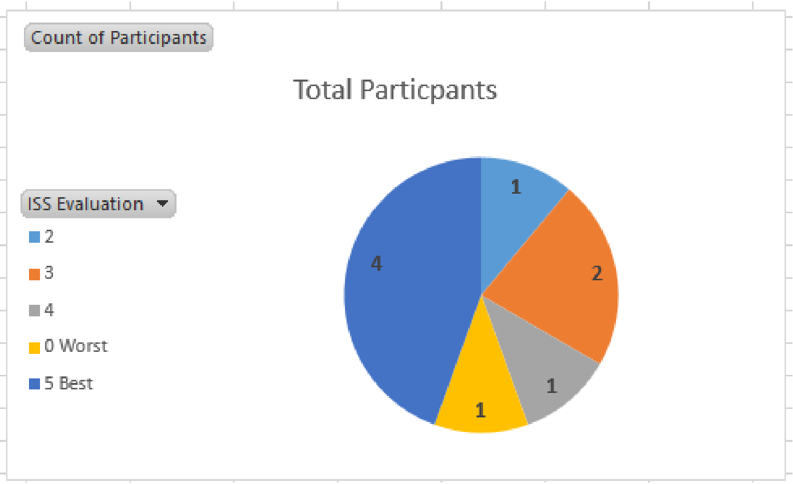


Figure 32 - Highest ranked participants by their ISS Role evaluations

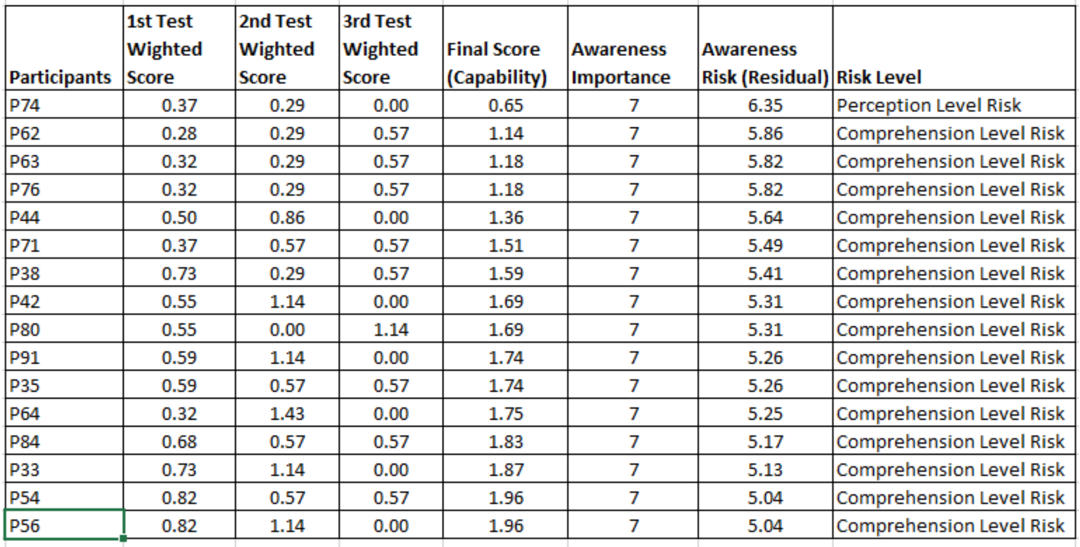
Finally, the majority of the 9 participants (8 out of 9) have an average of 7 days separating the two awareness sessions they had attended as opposed to greater periods of time the other participants had, which might have contributed to their low scores despite the fact that there was no correlational relationship found from the Regression test conducted on the average elapsed days between the two awareness sessions and the participants' scores on any of the tests.

In summary, it has been observed that the majority of the highest ranked 9 participants are from the Online Participant group, are at a young age (20-25) years old, have spent 5 years or more of online study, have an education level of a bachelor degree, have the best rating for their ISS roles and an average of 7 days period between the two awareness sessions attended.

5.24 Understanding the Attributes of Lowest Ranked Participants

On the other hand, the lowest ranked participants who achieved the lowest level of Phishing awareness and have the highest risk levels are those who scored between 0 and 2 out of 7 as far as their final score after summation of their scores from all the three tests is concerned. Running this criteria resulted in 16 participants being selected as opposed to only 9 selected as the highest ranked participants discussed in the previous section. These 16 participants make 16% of the whole sampled population. The lowest score recorded was 0.65 out of 7 with the highest risk level recorded as 6.35 whereas the highest score in the lowest score range (0-2) was 1.96 with a risk level of 5.04. 15 out of the 16 lowest ranked participants have a Comprehension Risk Level except for the last participant who has a Perception Risk Level. The Perception Level Risk ranges from 0 to 1, while the Comprehension Risk Level ranges from above 1 to 3. The first lowest ranked participant has the highest risk level in the Perception Level Risk range ever recorded which was 6.35, while the 16th (last) lowest ranked participant has the lowest Comprehension Level Risk ever recorded which was 5.04 was. Table 47 lists the 16 lowest ranked participants with their weighted scores and risk levels.

Table 47 - Lowest ranked 16 participants with their weighted scores and risk levels



Participants	1st Test Wighted Score	2nd Test Wighted Score	3rd Test Wighted Score	Final Score (Capability)	Awareness Importance	Awareness Risk (Residual)	Risk Level
P74	0.37	0.29	0.00	0.65	7	6.35	Perception Level Risk
P62	0.28	0.29	0.57	1.14	7	5.86	Comprehension Level Risk
P63	0.32	0.29	0.57	1.18	7	5.82	Comprehension Level Risk
P76	0.32	0.29	0.57	1.18	7	5.82	Comprehension Level Risk
P44	0.50	0.86	0.00	1.36	7	5.64	Comprehension Level Risk
P71	0.37	0.57	0.57	1.51	7	5.49	Comprehension Level Risk
P38	0.73	0.29	0.57	1.59	7	5.41	Comprehension Level Risk
P42	0.55	1.14	0.00	1.69	7	5.31	Comprehension Level Risk
P80	0.55	0.00	1.14	1.69	7	5.31	Comprehension Level Risk
P91	0.59	1.14	0.00	1.74	7	5.26	Comprehension Level Risk
P35	0.59	0.57	0.57	1.74	7	5.26	Comprehension Level Risk
P64	0.32	1.43	0.00	1.75	7	5.25	Comprehension Level Risk
P84	0.68	0.57	0.57	1.83	7	5.17	Comprehension Level Risk
P33	0.73	1.14	0.00	1.87	7	5.13	Comprehension Level Risk
P54	0.82	0.57	0.57	1.96	7	5.04	Comprehension Level Risk
P56	0.82	1.14	0.00	1.96	7	5.04	Comprehension Level Risk

It is observed that the 16 participants' scores on each test, do not demonstrate a consistent pattern of progress as some participants made gradual progress on the 1st and 2nd Test, but did not do well on the 3rd Test. Surprisingly, some participants got their best score on the 3rd Test, while having their lowest scores on the 1st and 2nd Test despite that the first two tests had less difficulty compared with the 3rd Test. Some other participants did not do any better as they went from one Test to another suggesting they did not have any benefits from attending the two awareness sessions.

Trying to understand if there was something special about these 16 participants by looking into their attributes, it has been found that 15 participants out of the 16 belong to the Face-to-Face participant group, while only one participant belongs to the Online participant group. This indicates that almost all the lowest ranked participants are from the Face-to-Face participant group. From an age perspective, all the 16 participants are between 20 to 25 years old which is the youngest age group among the 100 participant sample. From the education perspective, 14 participants out of the 16 are at a High School level, while only 2 have a bachelor degree. This indicates that the majority of the 16 participants have the lowest education levels as in Figure 33.

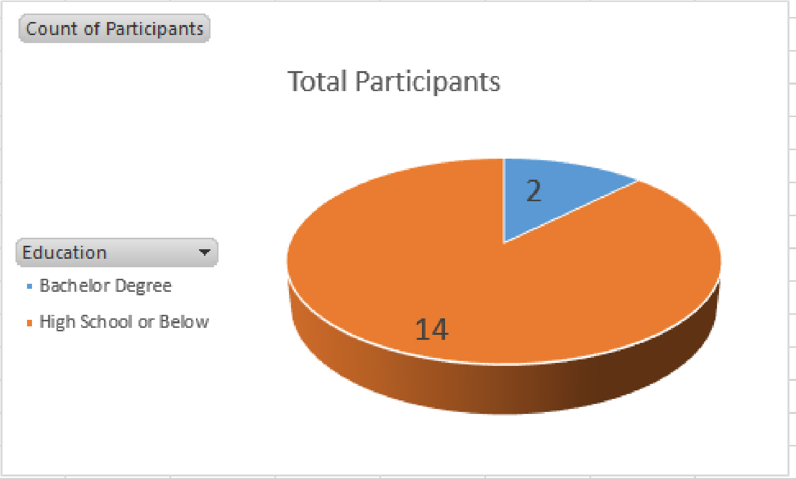


Figure 33 - Lowest ranked 16 participants by Education levels

As in Figure 34, almost half of the 16 participants (7 participants) have spent less than a year in online study, while 5 of them have spent 5 years or more and the remaining 4 have only spent 3 years in online study. This indicates no consistent relationship between the number of years of online study and the 16 participants' lowest scores. This also does not go in line with the results of the Regression tests conducted previously on all participants which found correlational relationships between participants' number of years in online study and the scores.

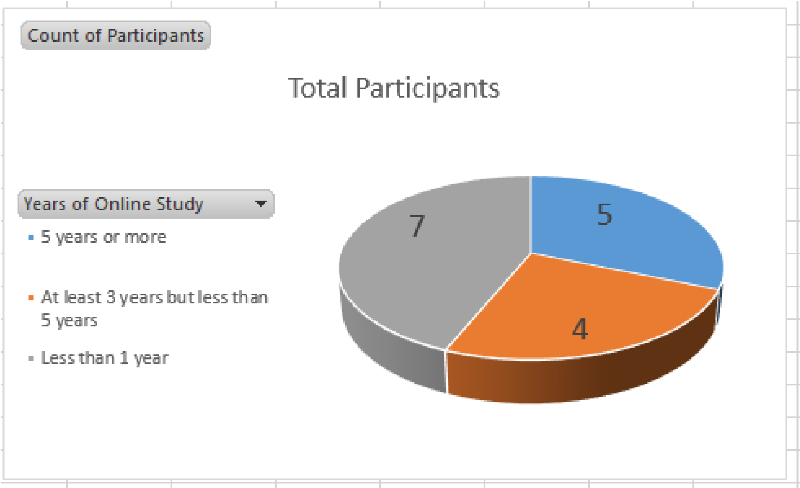


Figure 34 - Lowest ranked participants by Number of Years of Online Study

As Figure 35 shows, the majority of the lowest ranked participants (11 out of 16) evaluated the role of their ISS as the best, while 3 evaluated it as 4 on the scale of 5 and only 2 chose 'Not Sure'. Again, Figure 35 does not show any sensible relationship between the lowest ranked participants' evaluation rates of their ISS and their lowest scores. This goes in line with the results found from the Regression tests conducted previously on all participants which could not establish any correlational relationships between participants' ISS evaluation rates and their scores.

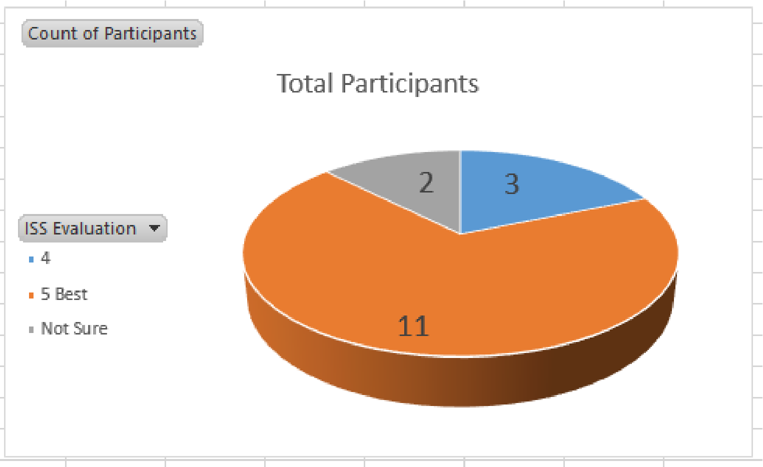
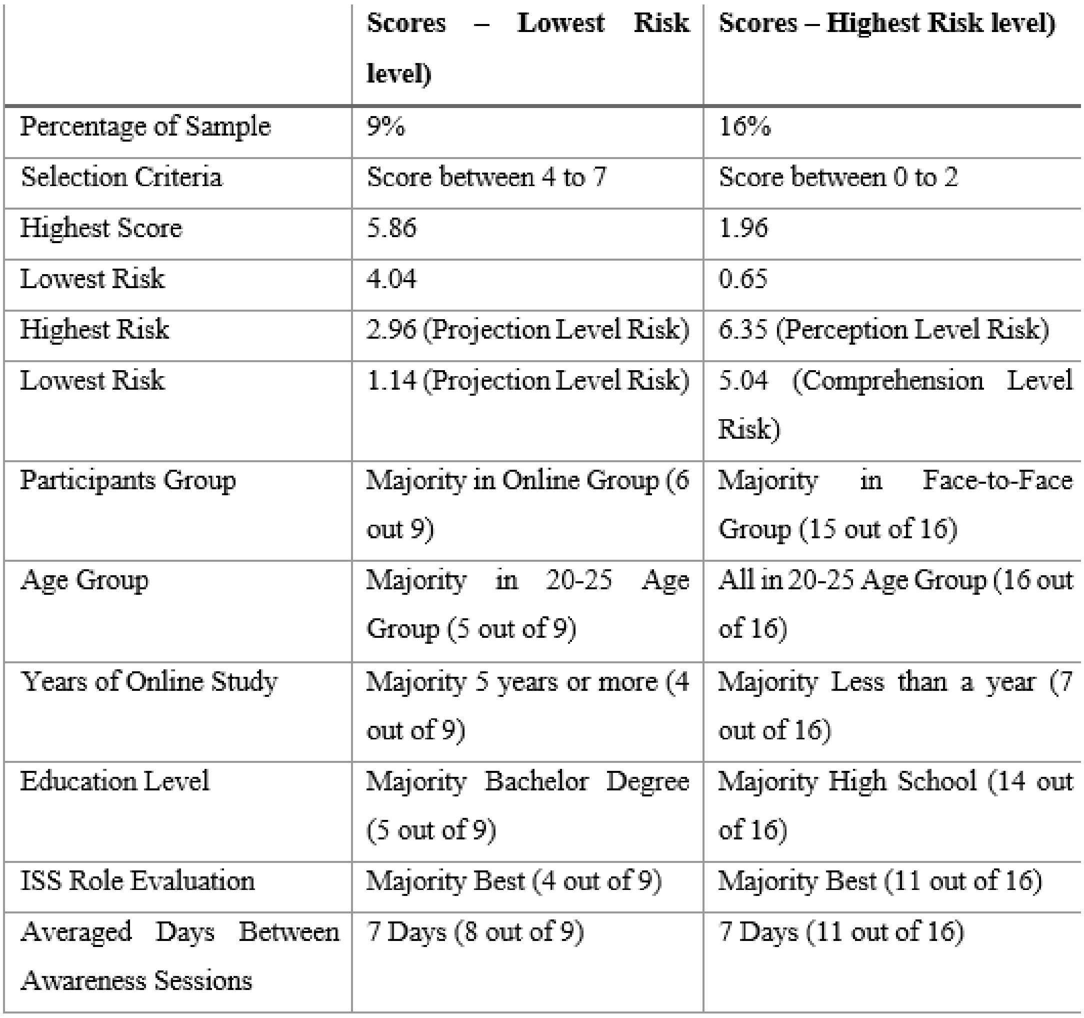


Figure 35 - Lowest ranked participants by their evaluation of ISS role

Finally, the majority of the 16 (11 out of 16) have an average of 7 days separating the two awareness sessions they had attended as opposed to only 8 out 9 of the highest ranked participants who had the same period of time between awareness sessions. There is no correlational relationships found as a result of conducting the Regression Test on the average elapsed days between the two awareness sessions and the participants' scores.

In summary, Table 48 contrasts and summarizes the results of comparing the highest and lowest ranked participants' attributes in terms of their scores.

Table 48 - Compaison between Highest ranked and Lower ranked participants



	Scores – Lowest Risk level)	**Scores – Highest Risk level)**
Percentage of Sample	9%	16%
Selection Criteria	Score between 4 to 7	Score between 0 to 2
Highest Score	5.86	1.96
Lowest Risk	4.04	0.65
Highest Risk	2.96 (Projection Level Risk)	6.35 (Perception Level Risk)
Lowest Risk	1.14 (Projection Level Risk)	5.04 (Comprehension Level Risk)
Participants Group	Majority in Online Group (6 out 9)	Majority in Face-to-Face Group (15 out of 16)
Age Group	Majority in 20-25 Age Group (5 out of 9)	All in 20-25 Age Group (16 out of 16)
Years of Online Study	Majority 5 years or more (4 out of 9)	Majority Less than a year (7 out of 16)
Education Level	Majority Bachelor Degree (5 out of 9)	Majority High School (14 out of 16)
ISS Role Evaluation	Majority Best (4 out of 9)	Majority Best (11 out of 16)
Averaged Days Between Awareness Sessions	7 Days (8 out of 9)	7 Days (11 out of 16)

In the next chapter, these results will be discussed in depth.

Chapter 6 Results Discussion

This research has aimed to achieve three goals. Firstly, to better understand the vulnerable behaviour factors and their exploitation in Phishing attacks. Secondly, to propose a conceptual framework to enhance awareness in online learners about these behavioural factors and their exploitation in Phishing attacks. Thirdly, to assess the effectiveness of the proposed conceptual framework by measuring the learning outcomes in online learners after applying the conceptual framework model. These three goals were fulfilled by answering the following 2 research questions and their 2 sub questions:

1. What are the vulnerable behavioural factors that are exploitable in Phishing attacks?
    a. How are these behavioural factors exploited in Phishing attacks?
2. What is the preliminary awareness level in online learners about these Phishing related behavioural factors?
    a. How can awareness about these behavioural factors and their exploitation in Phishing attacks be improved for online learners?

Thus, answering the 1st research question helped fulfil the first goal, while answering the 2nd research questions helped fulfil the second and third goals. The next sections will discuss how the conducted research contributed to answering each question and how that contributed to fulfilling the research goals.

6.1 Answering Research's 1st Question and Fulfilling Research's 1st Goal

One of the major threats to cyber security today is people's lack of awareness because people can be the weakest link in a security countermeasure systems, if they lack adequate awareness about information security risks such as Phishing (Gupta, Sharman, Raj, and IGI Global, 2009; Watson, Mason, Andrew, and Ackroyd, 2014); rendering other technical controls less effective if not properly addressed. Therefore, in order to effectively raise awareness about Phishing attacks, the reasons behind victims falling prey to phishing attacks repeatedly, despite all the awareness efforts being exerted, needs to be uncovered and explored to enhance the way awareness raising is conducted. To uncover these reasons, a literature review has been conducted to look through previous work analysing phishing attacks. The literature review revealed (see Chapter 2) that vulnerable behavioural factors were behind phishing attacks. The literature review resulted in the identification of 16 behavioural factors that were revealed by other research (see Section 2.6) to have been potentially manipulated in Phishing attacks exploiting human vulnerable behaviours and lack of awareness to lure unsuspecting victims. The literature review also revealed how these 16 behavioural factors were exploited to deceive victims in various phishing scenarios.

This research builds on the contributions of previous research such as Jagatic et al (2007), Kirlappos and Sasse (2012), Ibrahim (2016), Williams, Beardmore and Joinson (2017) and many others to identify and build a comprehensive list of potential Phishing vulnerable behavioural factors. No research was found to have gathered and compiled a comprehensive list of potential behavioural factors exploitable in Phishing attacks, before this research. Therefore, this research took on that duty and gathered 16 behavioural factors identified from the contributions of previous research. These 16 behavioural factors may not be considered inclusive of every possible behavioural factor exploitable out there, but it is a novel list. The identification of the 16 exploitable behavioural factors and the methods of exploiting them in Phishing attacks has also resulted in developing the proposed conceptual framework SAMFP to enhance awareness raising about these 16 behavioural factors which encapsulates the knowledge contributed from the literature review. Therefore by identifying the 16 behavioural factors and better understanding their exploitation in Phishing attacks from previous research, the first research question was answered and the first research goal fulfilled.

6.2 Answering Research's 2nd Question and Fulfilling Research's 2nd and 3rd Goals

In order to achieve the research's 2nd goal which is to improve awareness in online learners about the Phishing related behavioural factors and how they are exploited, the second research question had to be first answered by determining the preliminary level of awareness in online learners (participants in this research) about these behavioural factors. Knowing the participants' preliminary level of awareness is a necessary step towards improving their awareness through identifying the gaps and designing the appropriate awareness materials to address these gaps; which answers the second part of the research's 2nd question. The identified snapshot of the preliminary awareness in online learners sets the starting point for further awareness endeavours to take that awareness to the next level on Endsely's Situation Awareness Conceptual model (Endsley, 2015). Thus, the 1st Test was designed to explore participants' preliminary level of understanding about Phishing in general and their interpretations of why victims would fall prey to Phishing attacks. The 1st Test's results revealed that most of the participants (87/100) recognized the non-phishing scenarios more than they could recognize the Phishing scenarios. The average score recorded was 4.31/7 which revealed a level of moderate awareness about Phishing in general. However, the analysis of participants' reasons for Phishing revealed a considerably lower level of awareness compared to the average score recorded for recognizing phishing scenarios in the first place. The only behavioural factors referenced in participants comments as their reasons for considering these scenarios to be phishing were 'Temptation' and 'Urgency'. These 2 factors out of the 16 behavioural factors were referenced by 38% and 6% of the 100 participants respectively. The results also revealed participants' oblivion and uncertain knowledge of the other 14 behavioural factors as indicated by the comment 'No Specific Reasons' (9.63%) which was the most highly mentioned comment among all participants' answers. Also, generality of the participants' preliminary awareness about Phishing was indicated by the second most highly mentioned comments, 'Suspicious Request for Info' (9.53%), 'Unknown/Untrusted Sender/Caller' (6.80%) and 'Suspicious Course of Action' (5.09%), expressing general suspicion in each scenario without being able to identify the specific behavioural factors actually exploited in the scenario. A full discussion of the 1st Test analysis results is in Chapter 5, sections 5.1, 5.2 and 5.3.

The evaluation of the preliminary awareness in the 100 participants provided the answer to the research's 2nd question as well as fulfilling the research's 2nd goal. The value of providing information about the preliminary awareness level in online learners is very important for Information Security practitioners, who are responsible for raising employees' awareness about Phishing, to take this information and use it as a control group to contrast with their experimental groups and measure the progress of the learning outcomes and build their cyber security awareness efforts upon.

The analysis of the participants' scores and their responses to the 1st Tests' scenarios contributed a novel piece of knowledge about the preliminary awareness level in the participants about the 16 exploitable behavioural factors in Phishing. According to the literature reviewed, several efforts have undertaken such research to identify the preliminary level of awareness about Phishing in their subjects such as the work of Jagatic et al (2007), Kirlappos and Sasse (2012), Ibrahim (2016), Williams, Beardmore and Joinson (2017) and many others referred to in Chapter 2, Section 2.6. However, the 16 behavioural factors were not collectively covered by any of these efforts. In addition, this research has required its subjects to be online learners only in order to participate; which gives the knowledge obtained a novel value being focused on online learners in particular.

Gaining this knowledge about the participants' preliminary awareness level paved the road to the development of the SAMFP model to enhance the participants' preliminary awareness to higher levels of understanding about these 16 behavioural factors and their exploitation in Phishing attacks. The developed SAMFP model utilizes the knowledge gained about the preliminary awareness level in the participants and builds its awareness program upon that knowledge to help participants improve their awareness gradually through the phases of the model. In other words, the SAMFP model encapsulates the answer to the second part of the research's 2nd question and fulfils the research's 2nd and 3rd goals. Thus, the SAMFP is a novel model that incorporates three components by integrating Endsely's Situation Awareness model (2015) with a set of pedagogical guidelines recommended by Chen, Shaw and Yang (2006) to provide awareness and incorporating Poepjes' (2012) Information Security Awareness and Capability Model (ISACM) to measure the learning outcomes. During the conducted literature review, no research was found to have designed a similar framework like the SAMFP model that incorporates these three components together to provide similar features. The SAMFP model is unique in being a sustainable, continual and interactivity enabler in both system to human and human to human fashions. As such, the novelty of the SAMFP model lies in its carefully selected components and design which provide the following features: A) Sustainability of awareness delivery and continuity of learning through its two dynamic variables namely 'Time' and 'Space'. With Time, it provides awareness through iterative sessions across time until the desired level of awareness is achieved. Through 'Space', it provides flexibility and dynamicity of awareness delivery by facilitating awareness through online and Face-to-Face settings. B) Learning through interactive and collaborative methods using group discussions and role play. C) Measurability through continuous assessments that quantitatively measure awareness improvements and identify awareness gaps and address them through the iterative awareness sessions. The SAMFP model is discussed in detail in Chapter 3.

In the literature review, there are many cyber-security awareness systems and tools that follow different approaches to cyber security training. Some of these approaches rely on games for educating users about the risks of Phishing such as ‘Anti-Phish Phil’ (Kumaraguru, 2009), while other approaches depend on simulated phishing messages such as PhishMe (PhishMe.com). On the other hand, Phishing IQ tests is another approach to Phishing education (Anandpara et al, 2007). All these approaches have been successful to some extent in raising users’ awareness, but lacked some of the requirements of dynamic participant and trainer engagement, and interactive learning (Kumaraguru, 2009). These tools rely heavily on the system for providing interactivity. However, human collaborative interactivity through e.g. group discussions or workshops are missing in all these systems. This is where the SAMFP model incorporates both types of interactivities. The SAMFP model provides multiple scenario-based tests where users; in the first test, primarily use their imagination and intuition to recognize Phishing scenarios, while further iterative tests intervened by iterative awareness sessions extend users’ knowledge beyond recognition of Phishing to better understand the behavioural factors exploited in the Phishing scenarios. The SAMFP Model accomplishes this through interactive group discussions that involve role play among participants. Most of simulated message tools, game-based tools and Phishing IQs do not extend learning beyond Phishing recognition, whereas the SAMFP model extends knowledge to develop in-depth understanding of Phishing’s internal mechanisms and tactics. Nonetheless, some of these tools focus in their training on one or two of the behavioural factors such as trustworthiness, but do not focus collectively on as many as 16 behavioural factors nor cover them in depth as has been done in this research. They focus more on visual cues such as suspicious e-mail links (Kirlappos and Sasse, 2012). Training approaches and Phishing awareness tools are covered in more depth in Chapter 2. Section 2.5.

While the development of the SAMFP model encapsulates the answer to the research's question of how to improve awareness in online learners, the implementation of the model has provided the answer to the second part of the question which is how effective the model was in enhancing the participants' awareness about the 16 Phishing behavioural factors and hence fulfils the third goal of the research. The analysis of the three tests' scores (discussed in detail in Chapter 5) and its results (discussed later in the following sections) have shown that the level of awareness in the 100 participants who took the three tests and attended the two awareness sessions has significantly improved. Thus, the final results of the data analysis after all 100 participants completed the three assessment tests and attended the two awareness sessions, showed that the SAMFP model successfully helped 58% of the participants improve their awareness from the preliminary 'Perception' level to reach the 'Comprehension' level, and 41% to reach the 'Projection' level; which is the highest awareness level in the SAMFP model. The only exception to these achievements of the SAMFP model was one participant only who remained at the preliminary 'Perception' level. These gradual awareness improvements achieved meet the expected learning outcomes of the SAMFP model considering its cyclic nature spanning different periods of time and space. This means further improvements are potentially expected as more iterations of awareness sessions are conducted.

Moreover, the interactive delivery of the awareness sessions incorporating collaborative group discussion and role play have positively impacted the learning outcomes of the SAMFP model. The results of the data analysis on the SAMFP model's Time and Space variables showed a positive impact on the learning outcomes. With regards to Time, although the time periods separating the awareness sessions did not have any significant impact on the scores as far as the hypotheses tests are concerned, the results analysis showed that those who excelled in acquiring the highest awareness levels whether in the Online Group or the Face-to-Face Group all had lesser periods of time separating the conducted awareness sessions, compared to the longer periods the other participants had. This finding may indicate that the Time variable of the SAMFP model positively impacts the learning outcomes of participants as time periods decrease between sessions. With regards to the Space variable, although the Online participant group outdid the Face-to-Face participants in achieving higher awareness levels, the results did not indicate any significant impact on the learning outcomes with regards to the delivery method of the awareness sessions used. The results rather indicate that the excellent achievements of the Online participant group over the Face-to-Face participant group were attributed to other demographical factors such as age, education level and length of online experience. However, considering the limitations of the Online delivery method versus the Face-to-Face method (discussed in Chapter 3), it can be argued that the Face-to-Face setup could have a better position on the learning outcomes. Nevertheless, this argument still needs more confirmation through future research.

Next, the individual awareness levels (Perception, Comprehension and Projection) will be discussed in light of the analysis results to highlight awareness improvements. Also, these awareness improvements will be discussed in terms of the actual and weighted scores (hypotheses tests analysis). A discussion of the participants' awareness risk gaps and residuals will also follow. The highest ranked participants who achieved the highest awareness levels as opposed to those who only achieved the lowest awareness levels and the reasons that distinguished them from others will also be discussed. Finally, the research contributions and opportunities for future research will be discussed too.

6.3 Participant’s Preliminary Awareness about Phishing - Perception

The main objective of the 1st Test was to assess the preliminary awareness level namely ‘Perception’ about Phishing in the participants in general by asking them to identify if a certain situation stated in the test was a Phishing scenario or not. Also, if they recognized a scenario as Phishing, they were asked to tell why they thought it was. This was the way to implicitly evaluate their ability, to sense the underlying bait in the Phishing scenario by only relying upon their preliminary knowledge, and whether their explanation matched the actual behavioural factor used in the test scenario. The analysis of the results of the 1st Test revealed that most participants (up to 93%) were able to identify the non-phishing scenarios more than they could identify the phishing scenarios. This result can reassure to a certain extent that the majority of the participants (93%) had the competence to minimally distinguish a non-phishing scenario from a phishing scenario. However, this finding does not give the same assurance that they are as competent to recognize a phishing scenario as to identify a non-phishing scenario. This is due to not being adequately aware of the behavioural factors exploited by professional phishers.

On the other hand, the results also revealed that the most highly recognized phishing scenario was the one employing the ‘Over-confidence and Self-conscious’ behavioural factor as its implicit bait. This scenario was recognized by 79% of the participants, which was the highest number for any of the 16 phishing scenarios. However, this does not mean that the 97% who were able to recognize the phishing scenario were also able to identify the underlying ‘Over-confidence and Self-Conscious’ behaviour factor exploited by it. This means that they were led by their intuition of Phishing, not by their knowledge of these two behavioural factors to recognize the scenario as phishing. This was confirmed by the varying explanations and reasons the participants gave which in most cases were very general and high level. However, that does not belittle the phishing implications of the ‘Over-confidence and Self-Conscious’ behavioural factor in the scenario and its impact which aroused suspicion in participants and made them recognize the phishing nature of the scenario even though their expressions of the reasons were not a specific description of the factor as the behavioural factor used. This is due to their lack of awareness about these behavioural factors. Most of the participants’ comments about the phishing scenarios expressed general suspicion such as ‘suspicious request for information’ or ‘unknown sender/requestor’ and the like. These general expressions by the participants explained their preliminary perception of Phishing which does not go beyond their aroused suspicion of false requests for information from unknown/untrusted senders. These two general explanations were entered in 11 scenarios out of the 16 phishing scenarios recognized in the test and were mentioned 302 times across the 11 scenarios. It was also observed that the ‘No Specific Reason’ comment was also mentioned 91 times in 8 scenarios out of 16 phishing scenarios expressing participants’ uncertainty of any specific reason for believing these 8 scenarios were phishing; which again confirmed the participants’ perception of phishing aroused by the underlying behavioural factors without being able to specifically recognize them. Temptation was the fourth highest ranked comment out of 10 participants’ comments that were mentioned 87 times to explain why 5 scenarios out of the 16 were believed to be phishing.

Giving unspecific and general explanations (comments) to the phishing scenarios was not always the case. For example, in scenario #1 where temptation was the underlying behavioural factor, 38% of the 66 participants who correctly recognized this scenario as phishing, also commented that 'temptation' was the reason they believed was behind the phishing in the scenario; which was a correct match that described the exact behavioural factor underlying the phishing scenario. On the other hand, more generality was even expressed by participants using phrases such as 'Suspicious Links' and 'Suspicious Course of Action' to justify their judgement about certain scenarios of phishing. These two comments had much lower frequencies compared to the other comments explained already. The two comments were mentioned 37 and 28 times in 4 and 3 phishing scenarios respectively. One of the least frequently mentioned participants' comments was 'Urgency' which was mentioned for 15 times in two scenarios; which also correctly matched the same actual behavioural factor used in one of these two scenarios (Scenario #2) by 6% of the participants. Another comment with even lower frequency was 'Monetary Matters' which denoted fraud, also appeared 12 times in 1 scenario only.

It is concluded from the discussion above that Temptation and Urgency were the only behavioural factors out of the 16 factors that were correctly referenced by participants' comments with varying frequencies. Furthermore, the scenarios employing the behavioural factors: 'Convenience', 'Social Proof', 'Temptation' and 'Diffusion of Responsibility' received the next highest number of correct responses for being recognized as phishing by 74%, 71%, 66% and 66% participants respectively. On the other hand, the scenarios employing the behavioural factors: 'Over-Trust' and 'Reciprocation' received the lowest number of correct responses among all other 16 phishing scenarios. They were referenced by only 28% participants each. Therefore, it is observed that the behavioural factors that received the lowest number of correct answers indicate a significant low level of awareness in participants and hence may pose higher risks for participants. Therefore, more attention and awareness efforts should be dedicated to them to enhance the awareness about these less recognized behavioural factors and minimize the phishing risk they expose.

6.4 Participant’s Awareness Improvements about Phishing Behavioural Factors - Comprehension

The 1st awareness session was delivered to all the participants after their preliminary assessment test (1st Test) to explain to them the 16 behavioural factors and review with them how these factors were exploited in the phishing scenarios to lure victims. After they attended the 1st awareness session, they took the 2nd assessment test. The 2nd test manifests a higher level of difficulty as it requires the participants to identify the behavioural factors employed in a phishing scenario as opposed to the 1st Test which only required them to recognize if a scenario was phishing or not. Once again, the analysis of the 2nd Test’s results reflect upon the significance of the awareness improvements made as far as the actual scores are concerned by discussing the correct answers on the 2nd Test. For analysis of awareness improvements realized between the two tests, the results of the 18 hypotheses and the weighted scores are discussed next.

On the 2nd Test, the behavioural factors that were most highly identifiable by 85% of the participants were ‘Temptation’, ‘Urgency’ and ‘Curiosity’. On the other hand, ‘Diffusion of Responsibility’, ‘Social Proof’ and ‘Likability and Similarity’ were the least identifiable behavioural factors as they were only identified by 33% of the participants. Between the highest and the lowest of 85% and 33%, the second highest identifiable behavioural factors were ‘Authority’ and ‘Threatening, Fear and Anxiety’ which were identified by 81% of the participants. The third highest were ‘Overloading’ and ‘Commitment and Consistency’ which were identified by 62% of the participants; closely followed by ‘Reciprocation’ and ‘Interpersonal Relationship’ which were identified by 60% of the participants. ‘Over-trust’ and Convenience were identified by 54% of the participants compared to the ‘Over-Confidence’ and ‘Show-off’ which were identified by 46% only.

It is too early to judge whether these numbers of participants identifying the underlying behavioural factors on the 2nd assessment test indicate a significant improvement in their awareness levels before statistically testing the relative hypotheses. However, just by recalling the results from the 1st Test where the only two behavioural factors correctly referenced in the participants' comments were 'Temptation' and 'Urgency', a much bigger difference can be seen in the number of behavioural factors correctly identified by participants on the 2nd Test which increased by more than 50% most of the time. Again, the results of the 2nd Test confirm those from the 1st Test since 'Temptation' and 'Urgency' were the only behavioural factors that were correctly referenced by 66% and 46% of the participants respectively, the same behavioural factors are again part of the most highly identifiable factors (85%) on the 2nd Test. On the other hand, the behavioural factors: 'Over-Trust' and 'Reciprocation' which were the least identifiable on the 1st Test (28%) each, are now identifiable by more participants (54% and 60% respectively) on the 2nd Test. Although the phishing scenarios on the 1st Test employing the behavioural factors: 'Convenience', 'Social Proof' and 'Diffusion of Responsibility' received correct answers between 71% and 66% based on the number of participants recognizing them as Phishing, the same behavioural factors were less identifiable on the 2nd Test (4% and 33%) respectively.

Based on the analysis of the 2nd Test, a good level of understanding can be observed for most of the 16 behavioural factors especially those which were identified by more than 50% of the participants. The only exception which needs more attention in future awareness efforts is those behavioural factors identified by less than 50% of the participants namely 'Diffusion of Responsibility', 'Social Proof', 'Likability and Similarity', 'Over-Confidence' and 'Show-off'.

6.5 Participant’s Awareness Improvements about Phishing Behavioural Factors - Projection

The 3rd Test is set up the same as the 2nd Test, but with a higher level of difficulty that participants are required to not only identify the behavioural factors in 7 phishing scenarios, but also predict the behavioural factors most suitable for each scenario. Considering the increased difficulty on the 3rd Test, the percentages of the correct scores are not as high as those on the 2nd Test or the 1st Test. Hence, the weighted scores will be used instead to test for statistically significant awareness improvements by testing the relevant hypotheses later in this chapter.

Since ‘Over-Confidence’ and ‘Show-off’ were among the behavioural factors least identifiable by the participants on the 2nd Test, the same two behavioural factors received the highest correct answers on the 3rd assessment test after the participants attended the 2nd Awareness session and discussed these behavioural factors in more depth and demonstrated their use among their groups. ‘Over-Confidence’ and ‘Show-off’ were correctly identified by 56% of the participants compared to only 46% on the 2nd Test not taking into account the increased difficulty level on the 3rd Test at this stage. However, ‘Diffusion of Responsibility’, ‘Social Proof’ and ‘Likability and Similarity’ still are the least identifiable on the 3rd and 2nd tests by 33% and 4% of the participants respectively. It is notable that an accurate judgement cannot be made about the degradation of scores in these behavioural factors on the 3rd Test since the increased level of difficulty is not considered at this point of the analysis. With the exception of ‘Over-Confidence’ and ‘Show-off’ which were identified by 56% of the participants, there is a general decrease in scores in all other behavioural factors on the 3rd Test compared to the 2nd Test and that could be attributed to the multiple levels of difficulty incorporated in the 3rd Test. Thus, the second highest identifiable behavioural factors on the 3rd Test are ‘Overloading’ and ‘Commitment and Consistency’ which were identified by 45% of the participants compared to 62% on the 2nd Test. Nonetheless, ‘Temptation’, ‘Urgency’ and ‘Curiosity’ came in as the third highest identifiable factors on the 3rd Test by 23% of the participants compared to 85% on the 2nd Test. Similarly, ‘Reciprocation’ and ‘Interpersonal Relationship’ fell in fourth position, identified by 19% of the participants, whereas ‘Over-trust’ and ‘Convenience’ were identified by 14% on the 3rd Test as opposed to 60% and 54% of participants respectively on the 2nd Test.

Based on the above analysis, it is concluded that a general decrease in scores was observed due to the increased level of difficulty on the 3rd Test. However, the percentages of correct answers show a level of understanding that can easily be distinguished from those observed on the 1st and 2nd Tests respectively thanks to the two awareness sessions delivered to the participants before the 2nd and 3rd tests. Nonetheless, this is not a sound proof of awareness improvement. Yet, the need stands to discuss the analysis of the hypotheses to verify whether there was a significant awareness improvement that is statistically manifested. The utmost level of awareness achievable on the 3rd Test is the projection awareness level (Endsley, 2015) where most participants should be able to score full marks in each question. In light of this preliminary analysis of the 3rd Test's results, participants need more awareness sessions to improve their awareness levels and eventually their scores on the 3rd Test to reach the utmost level of awareness achievable. Special attention should be given to the 'Diffusion of Responsibility', 'Social Proof' and 'Likability and Similarity' behavioural factors which were the least identifiable by participants on both the 2nd and 3rd Tests and especially on the 3rd Test where they were correctly identified by only 4% of the participants.

6.6 Statistical Significance for Improved Awareness – Actual Scores

Before the discussion of the results of the 18 hypotheses, the actual (unweighted) scores achieved by participants on the three assessment tests will be compared and discussed. The objective of the actual scores comparison is to see how the actual scores are distributed on the scale of 7 among the three tests and how well participants did on each test.   This comparison will only look at the actual scores not taking into consideration the varying difficulty factor on each of the three tests; which means that a higher score on the 1st Test compared to a lower score on the 2nd Test or the 3rd Test may not necessarily mean that there were no awareness improvements achieved as the difficulty levels between the three tests are not the same. In the same way, a higher score on the 2nd or 3rd tests does not necessarily mean a significant awareness improvement either. Therefore, the weighted scores considering the different difficulty levels in each test will be compared instead when testing the 18 hypotheses that will tell us whether there was a statistically significant improvement or not. The result of the hypotheses tests will be discussed in detail soon.

The comparison between the actual scores reveals that participants achieved a higher mean score on the 1st Test (4.57) than on the 2nd Test (4.21) and 3rd Test (1.74) respectively. The difference in the mean scores between the 1st and 2nd Tests is not large, whereas it is so when compared with the 3rd Test's mean score. This explains the magnitude of the increasing level of difficulty incorporated in the hierarchical design of the three tests representing the three layered Endsley (2015) situation awareness levels. Similar differences are found between the medians, modes and standard deviations of the three tests. Further information on these aggregate statistics can be found in Chapter 5. Accordingly, on the 1st Test, the majority of participants (27%) scored 5 out of 7 whereas on the 2nd Test, the majority of participants (19%) scored as low as 4, and on the 3rd Test, the majority of participants (36%) scored as low as 2. In a similar manner, the lowest score recorded on the 1st Test was 2 out of 7 which was scored only by 2% of the participants, while the lowest score recorded on the 2nd and 3rd Tests was 0 which was scored by 2% and 11% of the participants on the 2nd and 3rd tests respectively. These decreasing scores also confirm the impact of the difficulty level increasing with each test.

6.7 Statistical Significance for Improved Awareness – Weighted Scores

The hypotheses are stated to test whether there was a significant improvement in the level of awareness achieved by the participants throughout the course of the three assessment tests and the two awareness sessions in between. The hypotheses will compare participants based on their weighted scores and groups namely, the online and the Face-to-Face groups and all participants (two groups combined). The results of the first 3 hypotheses tests (H1, H2, and H3) comparing the weighted scores of the two participant groups within each test revealed no significant awareness improvement manifested by either group over the other on any of the three assessment tests.

Nevertheless, the results of hypotheses (H4, H5 and H6) comparing the two participant groups' weighted scores between the three tests revealed that there were significant differences between the three tests for the Online and Face-to-Face groups and both groups combined as well. To find out which test of the three compared tests made that significant difference where a potential awareness improvement could have materialized, the next set of hypotheses verify that. Thus, the results of the hypotheses tests (H7, H8, H9, H13, H14 and 15H) comparing the weighted scores between the 1st Test and the 2nd Test pair, and then between the 1st Test and the 3rd Test pair, revealed significant awareness improvements achieved by both groups of participants on the 2nd and 3rd Tests respectively (2nd Test's weighted mean score = 1.18, 3rd Test's weighted mean score = 0.98) over the 1st Test (weighted mean score = 0.64) after attending the two awareness sessions between these tests. This finding confirms the findings from the actual scores comparison explained earlier.

However, hypothesis H10 comparing the weighted scores between the 2nd Test and the 3rd Test for all participants combined from the two participant groups revealed a significant difference in mean scores between the two tests. However, this significant difference is in the opposite direction of the expected learning outcomes to be gained after participants attended the 2nd Awareness Session where the higher mean should have been on the 3rd Test; not vice versa. The 2nd Test had the significantly higher mean score of (1.18) whereas the 3rd Test had the lower mean score of (0.98). Thus, the participants on the 3rd Test did not achieve significant awareness improvements over what had been achieved on the 2nd Test. This finding could be attributed to two reasons a) the multiplying difficulty level in the 3rd Test compared to the 2nd Test and b) the iterative nature of the applied SAMFP model where some participants might have required more iterations of awareness sessions to achieve the Projection level. Hypotheses H11 and H12 did not reveal any significance between the 3rd and 2nd Tests' scores for the Online and Face-to-Face participant groups.

6.8 Participants’ Demographics and their Impact on Awareness Improvement

The results of the analysis conducted on the participants’ demographical data and its correlational impact on the weighted scores of the three tests are reflected by the last 3 hypotheses (H16, H17 and H18). These 3 hypotheses tested for how much of the variations in the three tests’ scores can be attributed to the participants’ age, education levels and the number of years they spent in online study. In other words, the regression tests conducted on these three hypotheses helped determine which of these participants’ demographics could positively or negatively influence the participants’ scores on the three tests. As a result, it has been found that there are no significant relationships between these three independent variables and the participants’ scores on the three tests, except for one positive modest relationship between the participants’ number of years in online study and their scores on the 1st Test. This positive modest relationship with a coefficient of 0.078 can explain about 8% of the variations in the participants’ scores on the 1st Test. In other words, every higher category of ‘Years of Online Study’ can explain as much as 8% of the score.

Furthermore, the analysis of the participants' demographic attributes went a step further to test 2 more demographics for any correlational relationships with the scores on the three tests. These correlational tests considered 5 participants' attributes namely, 1) the age, 2) education level, 3) the number of years in online study, 4) average period between the two awareness sessions and 5) participants' evaluation rates of ISS. The objective of these correlational tests is to find out if any of these 5 participants' variables could have impacted the participants' scores on any of the three tests. These tests could yield a positive or negative correlation explaining that the affected score could either increase or decrease accordingly. Thus, the results of the conducted Spearman's Rank-Order Correlation tests revealed that age has a strong positive correlation with the scores on all the three tests indicating that the scores on the three tests would always increase for the participants who belong to older age groups (1st Test rs coefficient = 0.300; 2nd Test rs coefficient = 0.325; 3rd Test rs coefficient = 0.398). Similarly, education level also has a strong positive correlation with the participants' scores on the 2nd Test (rs coefficient = 0.322) and 3rd Test (rs coefficient = 0.314) only. In the same way, the Years of Online study variable has also a strong positive correlation with the scores on the 1st Test (rs coefficient = 0.477) and 3rd Test (rs coefficient = 0.245) only. On the other hand, the results revealed that the number of elapsed days (period) between the two awareness sessions and the participants' ratings of ISS have no correlational relationships with the scores on any of the three tests.

Based on the above, it can be concluded that the participants' age has a strong influence on the scores on all three tests as the older the participants were, the higher the scores they achieved. Also, the participants' education level has the same positive, significant influence on the scores, but only affecting the scores on the 2nd and 3rd Tests. The scores on the 1st Test were not affected by the education levels of participants, as the 1st Test was a preliminary level test that required no preparation from participants. In addition, this latter finding corresponds to the same results revealed by the regression test conducted on hypothesis (H16) for the 1st Test's scores and the effect of participants' education levels on the scores. Finally, the number of years participants had experienced online learning has also paid off on the 1st and 3rd Tests only, whereas, it had no effect on the 2nd Test's scores. These results support the results from the regression test conducted on hypothesis (H16) for the 1st Tests' scores where it revealed that the number of online study years had significantly explained the variations in the 1st Test's scores.

All participants in this research were recruited on the basis of having studied online for some time. However, there are other attributes of these participants that were collected during the 1st assessment Test that deserve attention and analysis to see if they had any correlational or influential relationship with the participants' scores. The participants' attributes which impacted the scores are the participants' age group, education level, the number of years in online study, while the average number of days elapsed between the two awareness sessions and the evaluation rating given by participants about their own Information Security Support (ISS) at their own establishments, whether educational or business organizations, did not have a significant impact on the scores. More information about participant's demographic data and their predefined categories can be found in Chapters 4 and 5.

Analysis of the participants' age groups revealed that (74%) of the participants are aged between 20 and 25 years old which indicates that the majority of participants are at a very young age or alternately said, are in the youngest age group defined for participants' selection. It is also observed that this high percentage (74%) covers a (100%) of the Face-to-Face participant group (60 participants), and (35%) of the Online participant group (40 participants). In addition, another 10% of the Online participants are aged between 26 and 30 years old which is the second youngest age group. This makes the percentage of the young participants aged between 20 and 30 years old as high as 84%. Thus, only 16% of the participants are distributed over the remaining 5 older age groups ranging from 31 to 51+ years old with each group spanning 5 years of age. This high percentage (84%) of participants being in the youngest age group supports the strong positive correlation found between the age groups and the Online participant Group's scores on all the three tests in which the older age group the participants were in, the higher the scores they tended to get.

Since the majority of the participants are in a young age group, they are also expected to have achieved an educational level that is as early as their young age would allow them to have achieved. Thus, it was found that the education level of 57 out of 60 participants in the Face-to-Face participant group is 'High School'; which goes in line with the finding that all the Face-to-Face participant group are aged between 20 and 25 years old. On the other hand, the education level of the majority of the Online participant group is a 'Bachelor Degree'. This is explained by the finding that the majority of the Online participant group are in older age groups as opposed to the Face-to-Face participant group who are all at the youngest age group. These results also conform to the findings from the correlational tests about the impact of the strong positive correlation between the education level and the scores where the higher an education level a participant had, the higher the scores they acquired on the 2nd and 3rd Tests.

Examining the participants' educational level and field of study revealed that there were 10 different disciplines and study fields. However, since the majority of participants were at the high school level, High School had the highest percentage of participants among all other fields which is 40%. The next highest were 'Oil and Gas' and 'Information Technology' with 17% and 13% of the participants respectively. The remaining 30% of the participants are scattered in fields such as Business Administration, Marketing, Fine Arts, HR, and Management etc. The fields and majors of study were not included in the correlational testing as the percentages are too widely dispersed and therefore would not produce meaningful results.

The number of years participants have spent in online study is also an important factor that influenced scores on the 1st and 3rd assessment tests as verified by the correlational tests. The statistics show that 54% of the participants have spent one year or less studying online. It was observed that the education level of more than half (57%) of those 54% participants was 'High School' and therefore they were also in the age group of 20-25 years old and made up the majority of the Face-to-Face participant group. The next highest category in the number of years of online study is those participants who have been studying online from 1 year up to 5+ years making 46% of all participants with more than half of them being trainees, whereas about 43% of these participants hold different jobs such as manager, senior management, team leader, regular employee, unemployed and others.

Taking these results and linking them to the results obtained from the correlational tests about the strong and positive relationship found between the number of years in online study and the scores on the assessment tests where the higher the number of years in online study, the better the chance participants could get a higher score.

With regards to the participants' evaluation of their ISS, 49% of them rated their ISS as 'Best' as opposed to only 2% whose rating for their ISS was 'Worst' on a Likert scale of 5. Almost two thirds (35 participants) of the 49% who rated their ISS as 'Best' were part of the Face-to-Face participant group, aged between 20 to 25 years old and hold a High School education level. The remaining 49% of the participants rated their ISS as 'Excellent' (21%), 'Very Good' (11%), 'Good' (7%) and finally 10% of the participants were neutral to evaluating their ISS choosing 'Not Sure' as their rating. Despite a high percentage of participants evaluating their ISS to have an excellent role on average in raising awareness about Phishing risks, the ISS evaluation ratings showed no significant correlations that would positively or negatively influence the participants' scores on any of the tests.

6.9 Participants' Awareness Risk Residuals

After discussion of the significant awareness improvements realized and the correlational relationships between the participants' demographics and the assessment tests' scores, the final part of the discussion covers how all of the above is interpreted into the SAMFP and Endsley's awareness levels (Endsley, 2015).

The participants are positioned relative to these awareness levels as far as their achieved awareness levels and the residual phishing risks. According to the findings obtained from the Information Security Awareness and Capability Model (ISACM) (Poepjes, 2012) that was applied to measure the residual risk of Phishing from the results realized after applying the proposed SAMFP model, it has been found that 58% of the participants have successfully achieved the 'Comprehension Level' on the SAMFP model which is considered the middle awareness level and therefore the risk residual is at the 'Comprehension Level Risk' category. Being at the Comprehension Level Risk category means that the participants are believed to be potentially protected from the lower level risk category namely (Perception Level Risk), but are still prone to the risks in their current awareness level category (Comprehension Level Risk) and beyond (Projection Level Risk). Nonetheless, 41% of the participants have also successfully achieved the highest level of awareness namely the Projection Level and therefore are believed to be exposed to the lowest risk residual which is the Projection Level Risk. On the other hand, only 1% which is equal to 1 participant only from all participants is still at the preliminary awareness level of Perception and is believed to be exposed to the highest risk residual category. This means that applying the proposed SAMFP has been successful in improving Phishing awareness for (99%) of all participants where it enabled more than half of the participants (58%) to improve their awareness from the preliminary (Perception Level) to the next level (Comprehension Level). Additionally (41%) of the participants achieved the highest level of awareness namely Projection. Hence, the risk of falling prey to Phishing attacks has been minimized in both awareness levels.

Achieving the highest level of awareness which means having the lowest risk residual does not mean full awareness or full protection as awareness is an iterative and dynamic process that has to constantly deal with the evolving Phishing risk. Achieving a certain awareness level e.g. the Perception level, Comprehension Level or the Projection level is relative to where participants are positioned within the range of their awareness risk levels. Since each level has its own range, some participants could just have achieved awareness to be at the beginning of the level, others between the middle and the maximum ranges of the level. This conclusion hereby provides an answer to the second research question of how awareness can be improved and sustained for online learners about Phishing behavioural factors.

Breaking down this conclusion into a more granular level to see its effect on the two participant groups namely the Online and the Face-to-Face groups, the analysis results tell us that 61% of the Face-to-Face group have achieved the Comprehension level as opposed to only 52% of the Online group. This means more participants in the Online Group have achieved the Projection level and therefore are exposed to less risk residual thanks to their higher awareness levels than in the Face-to-Face Group. Moreover the only participant (1%) who did not achieve higher than the preliminary Perception level belongs to the Face-to-Face Group.

Now digging into the results with more granularity to discuss the attributes of the highest ranked participants who achieved the highest levels of Projection during the course of this research. 9 participants were selected based on their scores falling in the range of the Projection Level. The first participant in the selection achieved the highest score of 5.86 out of 7. 6 out of the 9 participants have made gradual improvement manifested by their scores on the 1st Test through the 3rd Test. This indicates that the awareness improvements were gradual and in line with the proposed SAMFP awareness model as participants learned gradually from the awareness sessions they had attended. Six out of the 9 highest ranked participants belong in the Online Group as opposed to only 3 participants from the Face-to-Face Group.

From an age perspective, 7 out of the 9 highest ranked participants are aged between 20 and 35 years old, this indicates that they are at a younger age. Also, from the education perspective, 6 out of the 9 highest ranked participants (more than half) have a bachelor degree education level or above, this supports the conclusion from the statistical testing that there is a strong and positive correlation between the education levels and the participants' scores.

Finally, the number of years spent in online learning by the highest ranked 9 participants does not show a meaningful pattern as 4 of the 9 participants have been studying for 5+ years, while the (first) participant had been studying online for less than one year, while the others have spent different periods in online study. This inconsistency of this attribute within the selected data does not conform to the findings from the Regression and Correlational tests conducted on the whole dataset which showed a significant, positive and strong correlation between the number of years in online learning and the participants' scores. As far as the 9 highest ranked participants are concerned with their ISS evaluations, there is no consistent pattern found. This finding goes in line with the results from the Regression and correlational tests explained earlier. Last, but not least, it was observed that 8 out of the 9 highest ranked participants had an average of 7 days between attending the two awareness sessions as opposed to longer periods for other participants. However, there is no evidence from the regression or correlational tests that suggests any significance or indicates any relationship between the numbers of days elapsed and the participants' scores on any of the three assessment tests.

To summarize the findings from the analysis of the highest ranked 9 participants' demographical data, it is concluded that the majority of them are from the Online Participant group, are aged between 20 and 35 years old, have spent 5 years or more in online learning, have an education level of a bachelor degree, have an average rating of 'Excellent' for their ISS roles and an average of 7 days period between attending the two awareness sessions.

To complete this discussion, the attributes of the lowest ranked participants who are believed to be exposed to the highest risk residuals are discussed. 16 participants were selected based on the lowest scores in the range of 0 to 2 out of 7. The lowest score was 0.65 out of 7 with the highest risk residual recorded as 6.35 out of 7. The selection based on this range resulted in 16 participants. 15 out of the 16 participants have a Comprehension Risk Level except for one participant who has the Perception Risk Level.

Analysing the 16 participants' scores on each test reveals no consistent pattern of a learning curve as found with the highest ranked participants. Despite the fact that the 16 participants made some gradual progress on the 1st and 2nd Tests, they had an out-of-band downside pattern on the 3rd Test as they did not do as well as on the other tests. Yet, some of the 16 participants got their best scores on the 3rd Test compared to their lowest scores on the 1st and 2nd Tests, therefore might have benefited from attending the two awareness sessions before taking the 3rd Test. However, some of the 16 participants did not show any sign of improvement as they progressed from one test to the next indicating a steady pace despite attending the two awareness sessions.

Probing deeper in the 16 participants' demographic data looking for more clues that might explain the reasons for their low achievement, it was found that 15 out of the 16 participants were from the Face-to-Face group which might indicate that the Online Group gained better awareness than the Face-to-Face group. From an age perspective, all the 16 participants are aged between 20 to 25 years old which is the youngest age group as supported by the results of the correlational tests conducted on age groups. This shows that the younger the participants are, the lower the scores they tend to get. From the education perspective, 14 out of the 16 participants have a High School education level. The majority of the 16 participants have the lowest education levels, which gives an indication supporting the results from the correlational tests that the lower education level a participants have, the lower the scores they tend to get.

Finally, with regards to the number of years in online learning, the elapsed days between the two awareness sessions and the ISS evaluations, there were no consistent patterns or potential relationships found to justify the 16 participants not achieving higher levels of awareness. The findings from analysing the number of years in online learning for the 16 participants do not conform to the results of the regression and correlational tests conducted on the whole dataset. With regards to the average elapsed days between the two sessions, it was the only thing that was common between the lowest ranked and highest ranked participants which was 7 days. However, the results from the correlational and regression tests did not provide any evidence of a relationship that could impact the participants' scores positively or negatively.

6.10 Research Contributions and Future Research

The conducted research linking both realms Information Security and TEL have contributed the conceptual framework SAMFP which is based on a combination of learning theories to address the awareness residual risk in online learners in particular and general online users in general. As such, the outcome of the study could indirectly be applicable to any type of online users in general since all online users including e-learning users rely on the Internet and online environments which are venues for online Phishing attacks, to conduct their activities. Therefore, the contributions of this research to knowledge are the following:

1. **Identification of the 16 Phishing Related Behavioural Factors**: The identification of the 16 Phishing exploitable behavioural factors and the methods of their exploitation in Phishing attacks through the Literature review has provided the answer to the research's 1st question and also fulfilled the research's 1st goal by better understanding these behavioural factors.

2. **A snapshot of current online learners' awareness of Phishing behavioural factors**: Through the 100 participants' answers to the 1st assessment test, a snapshot of current awareness level in a sample of 100 online learners about Phishing behavioural factors has been obtained to answer the research's 2nd question (RQ2).

3. **A novel conceptual framework model called SAMFP:** The development and introduction of the novel conceptual framework model 'SAMFP' helped raise participants' awareness from preliminary awareness level (Perception) to higher levels e.g. 'Comprehension' and 'Projection'. This was a fulfilment of the research's 2nd goal.

4. **Evidence of the efficacy of SAMFP in raising awareness levels in online learners:** The implementation of the SAMFP model to enhance online learners' awareness about Phishing behavioural factors has enabled measuring awareness improvements and identifying awareness gaps which in turn helped in the assessment of the efficacy of the SAMFP model in raising awareness in online learners. This has provided answers to the research's 2nd question and helped fulfil the research's 2nd and 3rd goals.

Since the sampling approach taken in this research is non-random convenience sampling requiring participants only to be online learners regardless of any other traits such as specific location, discipline or culture, Therefore, the distribution of participants in terms of location and culture was not part of the selection strategy or the data analysis. Hence, this could also be a potential opportunity to extend this in future research studying the effect of a particular sampling distribution that is based on specific demographics such as culture, nationality, location, education major or career on online learners' susceptibility to Phishing attacks.

This research also opens other venues for future research. Since the focus of this research is online learners, further research could take the SAMFP model and apply it to test its efficiency for other categories of online users. For example, applying the SAMFP model on gender-specific online learners, specific age group, or specific specialities and compare the differences in the learning outcomes with other research experiments such as this one. Another opportunity for future research is to apply the model to a different class of users such as specific school students or a group of employees.

Furthermore, extending the cycle of implementing the SAMFP model to include more iterations of awareness sessions and a bigger number of assessment tests could be another future expansion to this research. Since awareness improvements in SAMFP are time and space dependent, extended experiments will test for more significance of the impact of the two dynamic variables of the SAMFP model, which can be built upon by future research.

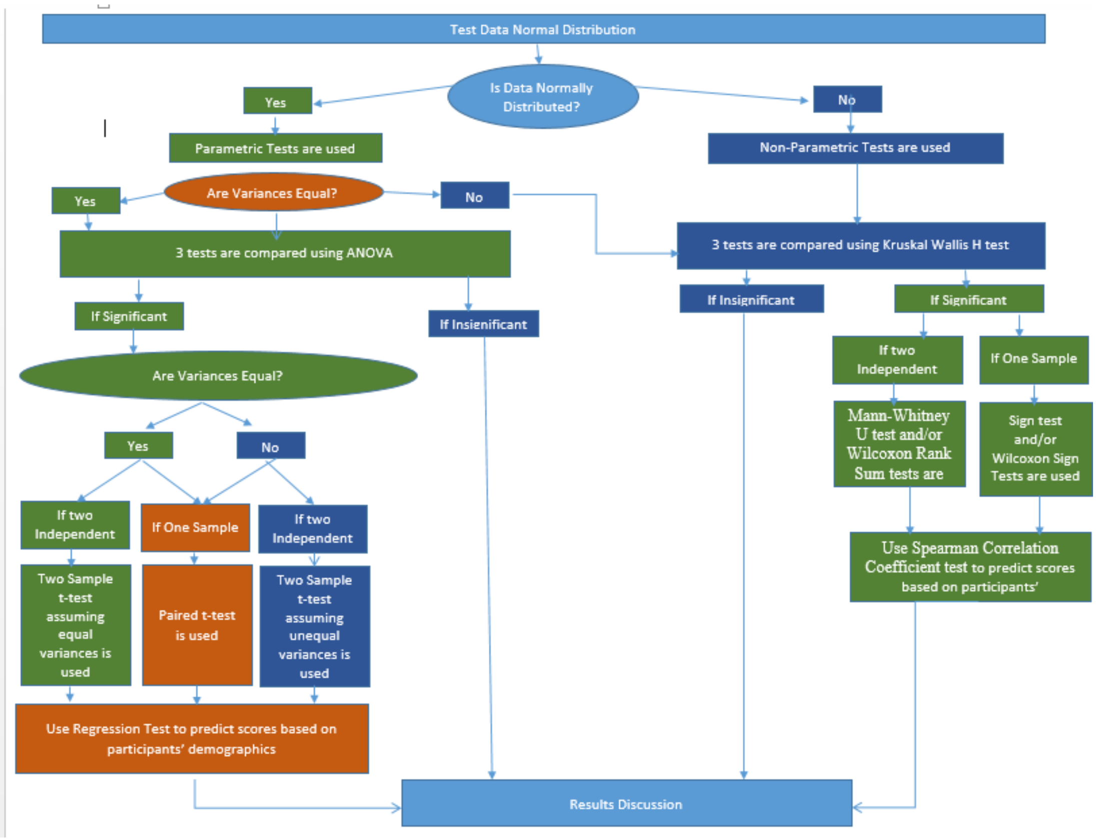
Finally, I would suggest to employ the SAMFP model within an organization that favourably and solely depends on technical controls as a silver bullet for protection against cyber-attacks. This will provide a great opportunity for information security practitioners and researchers to study how incorporating the SAMFP model as a second layer of defence within that organization could benefit the organization by enhancing its immunity and protection against Phishing attacks. This also goes in line with the principal of ‘Defence in depth’ which encourages designing security controls in layers where incorporating an effective cyber-security awareness model as an additional layer of defence can complement the technical controls already in place by greatly enhancing the level of the organization’s security posture.

Appendices

Appendix A - Statistical Testing

A.1 Statistical Testing Process

To start the statistical analysis, different attributes of the data such as data distribution (Bai and NG, 2005) and skewness are initially assessed and measured in order to determine the type of statistical tests appropriate to use on the data e.g. parametric or non-parametric tests. Parametric tests are used when the data is assumed to be normally distributed whereas the non-parametric tests are used when the assumption of normal distribution is not met (Cohen, Manion and Morrison, 2011). Therefore, the following test process is followed as in Figure 36.



Test Data Normal Distribution
Is Data Normally Distributed?
Yes
No
Parametric Tests are used
Non-Parametric Tests are used
Yes
Are Variances Equal?
No
3 tests are compared using ANOVA
3 tests are compared using Kruskal Wallis H test
If Significant
If Insignificant
If Insignificant
If Significant
If two Independent
If One Sample
Are Variances Equal?
Mann-Whitney U test and/or Wilcoxon Rank Sum tests are
Sign test and/or Wilcoxon Sign Tests are used
Yes
No
If two Independent
If One Sample
If two Independent
Use Spearman Correlation Coefficient test to predict scores based on participants'
Two Sample t-test assuming equal variances is used
Paired t-test is used
Two Sample t-test assuming unequal variances is used
Use Regression Test to predict scores based on participants' demographics
Results Discussion

Figure 36 - Statistical Testing Process in this research

The statistical analytical testing process above starts by initially testing the data (The Three tests' scores) individually for normality as normal distribution in the tested data is an assumption of some of the statistical tests (Shapiro and Wilk, 1965) used in the testing procedure which will be explained later. Data normal distribution is tested by (1) comparing the data to a normal probability plot (Ghasemi and Zahediasl, 2012; Bandyopadhyay n.d.; Filliben, 1975), (2) using Frequency Bins and Histograms (Sircombe, 2004; Scott, 1979' Tukey, 1977) to graphically draw the normal distribution curve of the data and (3) finally measuring the data skewness significance (Altman and Bland, 1996). More on this will be discussed in detail in the data normality tests section later in this appendix.

The testing process will branch out into two different paths depending on the result of the data normality tests conducted. If the data is proven to be normally distributed, a set of parametric tests (Sheskin, 2003) will be used while if otherwise, a set of non-parametric tests will be used instead. The other assumption required by some parametric tests such as the Analysis of Variance (ANOVA) test (Ruxton, 2006; Febrero and Fraiman, 2004) is that compared datasets should have equal variances (Levene, 1960; Markowski and Markowski, 1990). Therefore, after data normality is tested, F-test (Bernhardson, 1975) is used to test for variance equality between any two datasets or more that will be consumed by ANOVA test which is the next step in the testing process. If the 3 dataset pairs (Scores of the 1st Test, 2nd Test, and 3rd Test) are shown to have equal variances, all 3 tests' scores datasets are compared using ANOVA test to check if the differences in the mean scores between the three tests' scores are significantly different. However, the ANOVA test does not report which dataset is the source of the significance (Rice, 2017). Therefore, the next step in the process is to test each pair of datasets individually to identify where the significance is among the three datasets compared. There are two statistical tests that are appropriate for verifying the significance of difference between the mean scores of any two datasets compared. These tests are the One Sample Paired t-Test and the Two Independent Samples t-Test (Zimmerman, 1997; Knapp, 1978). The One Sample Paired t-Test is used when the two datasets (tests' scores) compared belong to the same sample e.g. the scores of the same participant group are compared between two different tests. For example, this type of t-Test is used to compare the scores of all the 100 participants, the Online participant group and the Face-to-Face participants' datasets individually between the three tests in test pairs e.g. the 1st Test and the 2nd Test pair, the 2nd Test and the 3rd Test pair and finally the 3rd Test and 1st Test pair. The second type of t-Test is the Two Sample Independent t-Test which is used to compare the scores of two different samples (different participant groups) between the three test pairs and within each test. For example, the Independent Sample t-Test is used to compare the mean scores of the Online participant group with those of the Face-to-Face participant group between the 3 test pairs and within each test. However, the Independent Samples t-Test comes with two flavours; one of which is designed to test two datasets assuming unequal variances whereas the other one is designed to test two datasets assuming equal variances. Thus, each pair of dataset's variance is tested for equality first. For example, the variance between the mean scores of the 1st Test and 2nd Test's scores dataset is calculated to determine the right type of test to use as illustrated in Figure 38.

On the other hand, if the data is not shown to be at an acceptable level of data normality, non-parametric tests (Siegel, 1956) can be used as these non-parametric tests do not require their underlying data to be normally distributed or to have equal variances (McKnight and Najab, 2010). Similar to the parametric tests path, the first test used in the non-parametric process path is to compare the mean scores of all the three tests for significance using Kruskal-Wallis H test (Wallace, 1995). If Kruskal-Wallis H test shows that the data demonstrates a significant difference between the compared mean scores, further non-parametric tests are to be used to identify which of the two datasets was the source of the significance. The choice of the non-parametric tests for testing significance between the 3 individual test pairs depends on the group of participants involved in each test. If the compared scores belong to the same group of participants (one sample) e.g. the Online participant group or the Face-to-Face participant group or the participants of the two groups combined, Sign Test (Dixon and Mood, 1946) and/or Wilcoxon Rank Sign Test (Woolson, 2008) are used. While if the scores compared belong to two independent groups of participants, Mann-Whitney U Test (McKnight and Najab, 2010) and/or Wilcoxon Rank Sum Test (Gibbons and Chakraborti, 2011; Conover and Iman, 1981; Siegel, 1956) are used.

Finally the last tests performed as part of the testing process are the Simple Regression Test (Pedhazur, 1997) and the Spearman Correlation Coefficient Test (Zar, 1972) used for the Parametric and the Non-parametric tests paths of the process respectively. These two tests are used to measure the impact of participants' attributes (independent variables) such as age, education, periods of online study, intervals between awareness sessions and role of ISS on the tests' scores (Knofczynski and Mundfrom, 2008) and to use the results from these tests to predict future tests' scores.

The following section will explain how the data analysis using the process above was conducted on the stated hypotheses explained in Chapter 5 sections 11-17.

A.2 Testing Data Normality and Variance Equality

The methods used for testing data normality and variance equality according to the testing process explained earlier and their results are described. This is the preliminary work to prepare the data for further testing and analysis against the 18 stated hypotheses. This data preparation includes normalizing and standardizing the scores first (explained earlier in Chapter 4 Sections 13-15) and then testing the data to see if it satisfies the assumptions of the statistical tests intended for use in the hypotheses analysis.

A.3 Data Normal Distribution Tests

The following normality tests are applied to all the 9 datasets explained earlier that provide all the data consumed by the statistical tests. This preliminary step is required to verify whether the statistical tests' assumptions about the data are met. In order to test the normal distribution attribute of the data, two methods are used. First, data skewness (Altman and Bland, 1996) was calculated and graphically compared to a normal probability plot (Ghasemi and Zahediasl, 2012; Bandyopadhyay n.d.; Filliben, 1975) drawn to illustrate the level of data skewness against the normal distribution curve of the data. Second and complementary to the first method, Histograms and Frequency Bins (Sircombe, 2004; Scott, 1979' Tukey, 1977) are developed to graphically demonstrate the curve of normal distributions built into the data. The objective of using multiple methods is to validate the output from one method against the other. These methods and their results are explained below.

Data Skewness and the Normal Probability Plot: The normal probability plot is a graphical way of looking at the data to verify its normal distribution by sight. This is accomplished by simply plotting the data on a scatterplot and comparing its plotted line against a normalized data plot (Bandyopadhyay n.d.; Filliben, 1975). To produce a normalized plotted line of scores against which the original scores are compared, two sets of scores namely the expected scores and the z-scores corresponding to every original score in the data are calculated based on the probability figure obtained by using the Cumulative Distribution Function (CDF) (Abramowitz and Stegun, 1967; Brownlee, 1965). Then, the original set of scores and the expected scores are both plotted and compared to the set of z scores; resulting in two plotted lines. The resulting two plotted lines are then compared. If the two plotted lines appear to coincide, then that indicates that the data is normally distributed, otherwise skewed.

Even if the data appears to be skewed, yet the significance level of skewness is evaluated to determine if the data skewness is within an acceptable range for consumption by a parametric test that requires normality (Mardia, 1970). The skewness significance range of the data is measured by comparing the calculated skewness value against twice the roughly estimated standard error of skewness which is obtained by the following formula: “2 multiplied by the square root of the number of original scores divided by 6”. This means that if skewness is within the range of twice the skewness standard error, it is not considered a significant normality issue (Trochim and Donnelly, 2006). If the measured skewness value is greater than the calculated result of the formula then, the data (original scores) are significantly skewed and hence cannot satisfy the normality requirement of the test (Shapiro and Wilk, 1965).

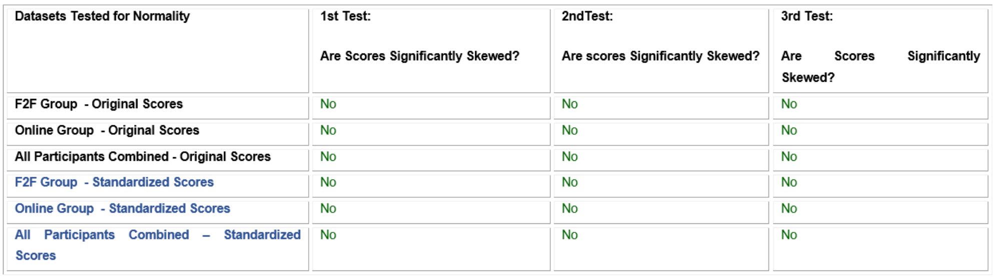
Below are the formulas used in the previous normality test that calculate the CDF, Expected Value of the scores, the z-value and the significance of skewness of the original scores. All the other formulas are dependent on the CDF value:

- CDF = the first score’s CDF value is calculated with the formula “1 / (2 * Count of Scores)”. Then the next score’s CDF value is generated by adding the previous CDF to the result of the formula “2 / (2 * Count of Scores)” and so on until the last score’s CDF value is generated.

- Expected Value = is calculated using the Excel Function “NORM.INVERSE (CDF , Mean of Scores , Standard Deviation of Scores)”
- Z-Value = is calculated using the Excel Function “NORM.S.INVERSE(CDF)”

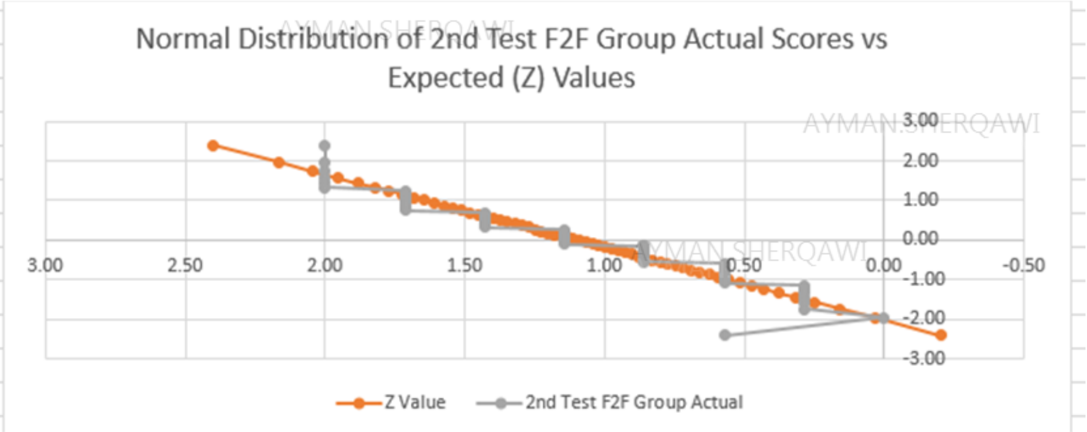
These normality tests are applied to the 9 datasets in two rounds; once using the original scores and another round of tests using the standardized scores. A standardized score is a more normalized form of the original score which is calculated by subtracting the original score from the mean score and dividing it by the standard deviation of original scores. It is a verification technique by which the results of conducting the normality tests on the original scores are validated against those of the standardized scores. As a result, both rounds of tests collectively indicated that all the 9 datasets (original scores and standardized scores) show an acceptable level of normality that allow them to be consumed by the intended statistical tests. Table 49 shows the collective normality tests results for both the original and standardized scores.

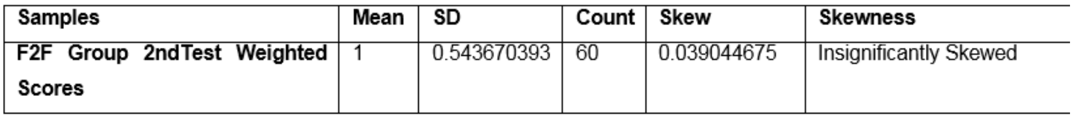
Table 49 - collective normality tests results for both the original and standardized scores for the three tests



Datasets Tested for Normality	1st Test: Are Scores Significantly Skewed?	2ndTest: Are scores Significantly Skewed?	3rd Test: Are Scores Significantly Skewed?
F2F Group - Original Scores	No	No	No
Online Group - Original Scores	No	No	No
All Participants Combined - Original Scores	No	No	No
F2F Group - Standardized Scores	No	No	No
Online Group - Standardized Scores	No	No	No
All Participants Combined – Standardized Scores	No	No	No

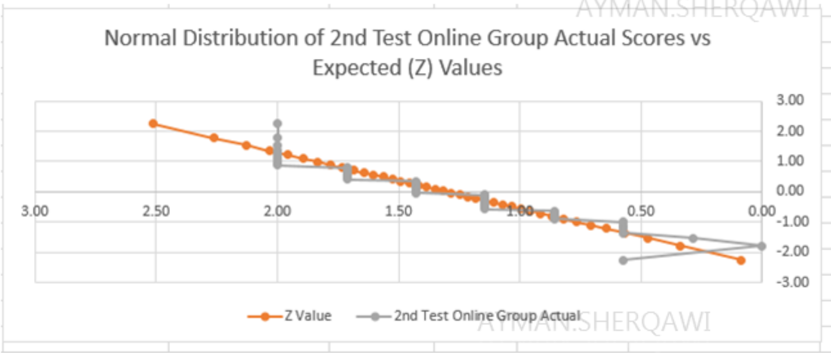
Moreover, the probability plot drawn for each dataset shows relatively acceptable level of normal distribution for the plotted scores against the normal z-scores’ plotted line (Ghasemi and Zahediasl, 2012) as illustrated in Figures 37 through 42 plotting the original scores of the Face-to-Face and Online Groups’ participants for the 1st 2nd and 3rd Tests respectively.

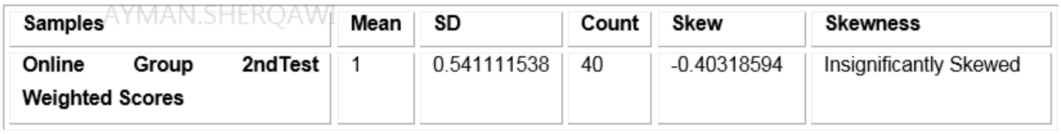




Samples	Mean	SD	Count	Skew	Skewness
F2F Group 2ndTest Weighted Scores	1	0.543670393	60	0.039044675	Insignificantly Skewed

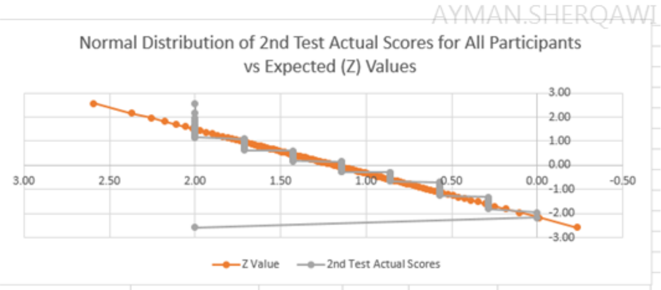
Figure 37 - Normal Distribution test for 2nd Test scores of the Face-to-Face Group using Probablity Plot





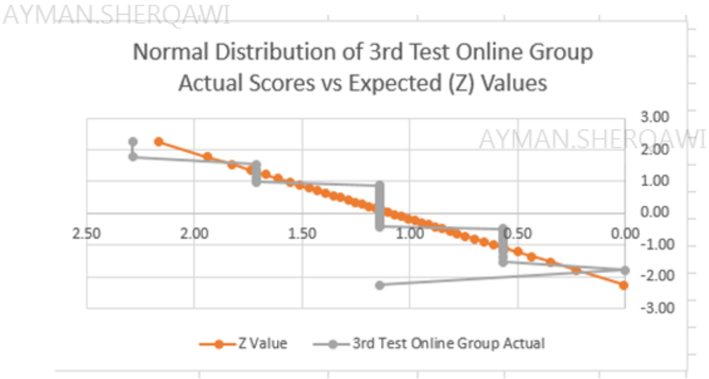
Samples	Mean	SD	Count	Skew	Skewness
Online Group 2ndTest Weighted Scores	1	0.541111538	40	-0.40318594	Insignificantly Skewed

Figure 38 - Normal Distribution test for 2nd Test scores of the Online Group using Probability Plot



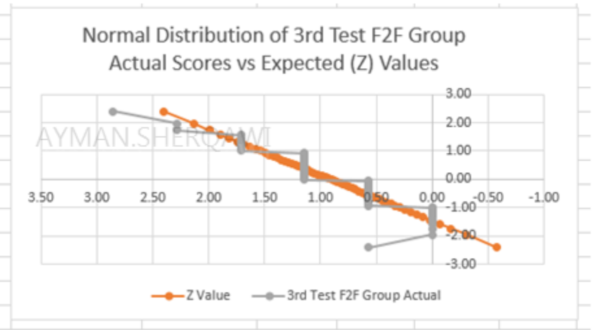
Samples	Mean	SD	Count	Skew	Skewness
2ndTest Weighted Scores – All Participants	1	0.548811778	100	-0.127512342	Insignificantly Skewed

Figure 39 - Normal Distribution test for 2nd Test scores of the all participants combined using Probability Plot



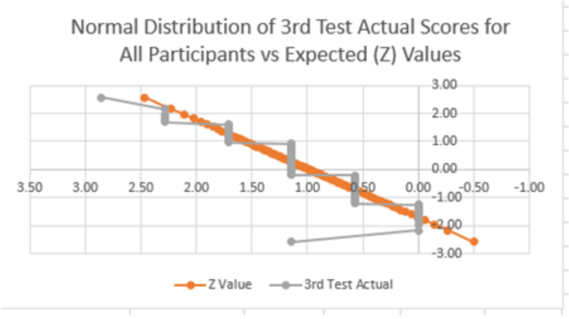
Samples	Mean	SD	Count	Skew	Skewness
Online Group3rd Test Weighted Scores	1	0.48068711	40	0.46786954	Insignificantly Skewed

Figure 40 - Normal Distribution test for 3rd Test scores of the Online Group using Probability Plot



Samples	Mean	SD	Count	Skew	Skewness
F2F Group 3rd Test Weighted Scores	1	0.62416665	60	0.62310154	Insignificantly Skewed

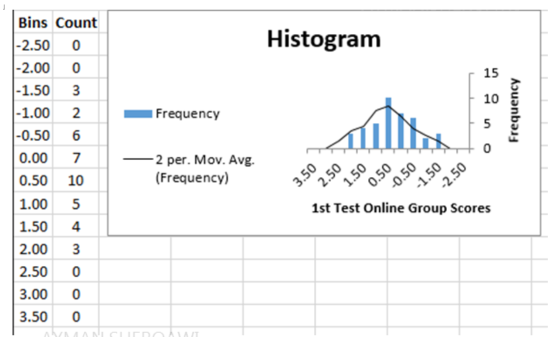
Figure 41 - Normal Distribution test for 3rd Test scores of the Face-to-Face Group using Probability Plot



Samples	Mean	SD	Count	Skew	Skewness
2nd Test Weighted Scores – All Participants	1	0.57473785	100	0.46925466	Insignificantly Skewed

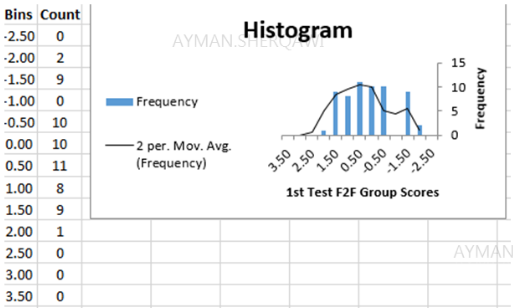
Figure 42 - Normal Distribution test for 3rd Test scores of the all participants combined using Probability Plot

Frequency Bins and Histograms: The second method to test and verify the normality of the data is to use the Frequency Bins and Histograms to graphically show the level of normal distribution curve of the data. This is to complement and support the results revealed by the normal probability plot and the measured skewness significance obtained from the previous tests. Frequency bins and histograms are a technique to visualize the distribution curve of the data by bins or buckets filled with the frequency (occurrences) of each score occurring at pre-set intervals (bins) over the distribution curve (Sircombe, 2004; Scott, 1979' Tukey, 1977). This test is conducted on all the 9 datasets with their original sores and standardized scores. The results from these histograms show that all the 9 datasets demonstrate a relatively similar level of normality with non-significant levels of skewness conforming to the results obtained by the previous normality plot tests. Following are some of the histograms and bins shown in Figures 43 through 53 for some of the participants' standardized scores on the 1st, 2nd and 3rd Tests respectively.



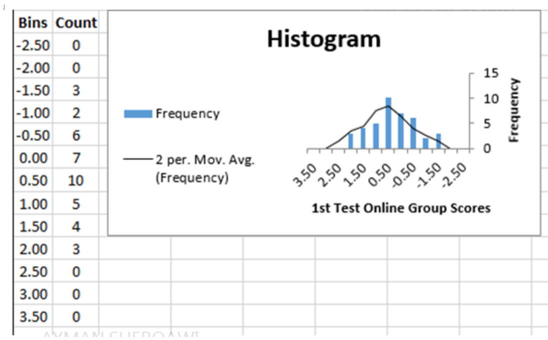
Bins	Count
-2.50	0
-2.00	0
-1.50	3
-1.00	2
-0.50	6
0.00	7
0.50	10
1.00	5
1.50	4
2.00	3
2.50	0
3.00	0
3.50	0

Figure 43 - Normal Distribution test for 1st Test scores of the Online Group using Histograms and Bins



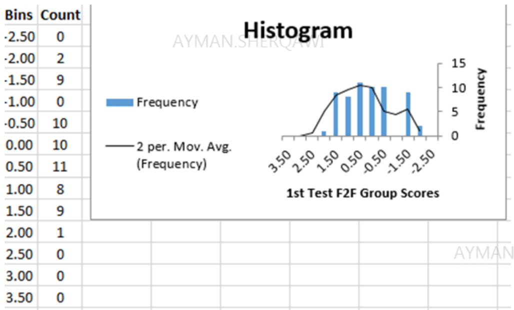
Bins	Count
-2.50	0
-2.00	2
-1.50	9
-1.00	0
-0.50	10
0.00	10
0.50	11
1.00	8
1.50	9
2.00	1
2.50	0
3.00	0
3.50	0

Figure 44 - Normal Distribution test for 1st Test scores of the Face-to-Face Group using Histograms and Bins



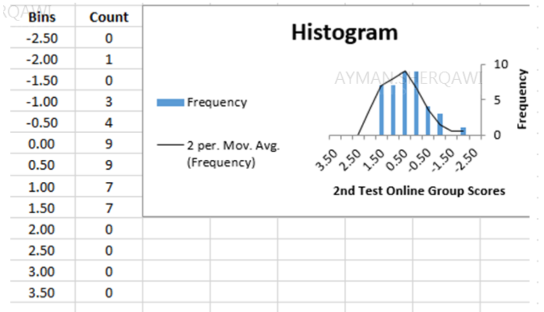
Bins	Count
-2.50	0
-2.00	0
-1.50	3
-1.00	2
-0.50	6
0.00	7
0.50	10
1.00	5
1.50	4
2.00	3
2.50	0
3.00	0
3.50	0

Figure 45 - Normal Distribution test for 1st Test scores of the Online Group using Histograms and Bins



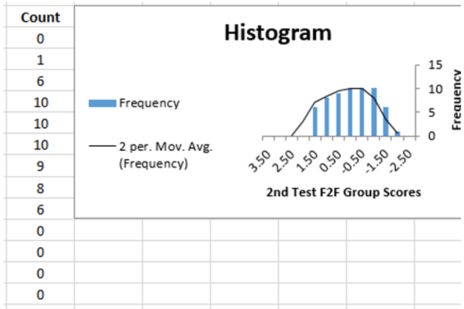
Bins	Count
-2.50	0
-2.00	2
-1.50	9
-1.00	0
-0.50	10
0.00	10
0.50	11
1.00	8
1.50	9
2.00	1
2.50	0
3.00	0
3.50	0

Figure 46 - Normal Distribution test for 1st Test scores of the Face-to-Face Group using Histograms and Bins



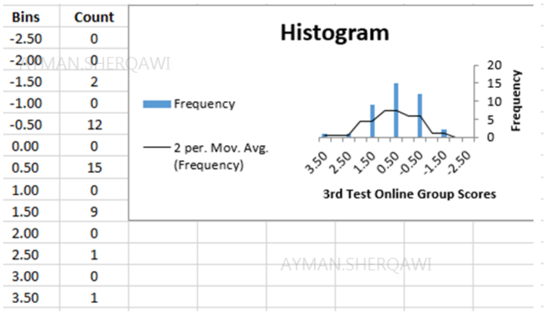
Bins	Count
-2.50	0
-2.00	1
-1.50	0
-1.00	3
-0.50	4
0.00	9
0.50	9
1.00	7
1.50	7
2.00	0
2.50	0
3.00	0
3.50	0

Figure 47 - Normal Distribution test for 2nd Test scores of the Online Group using Histograms and Bins



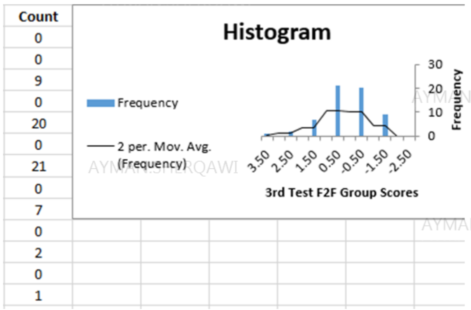
Count
0
1
6
10
10
10
9
8
6
0
0
0
0

Figure 48 - Normal Distribution test for 2nd Test scores of the Face-to-Face Group using Histograms and Bins



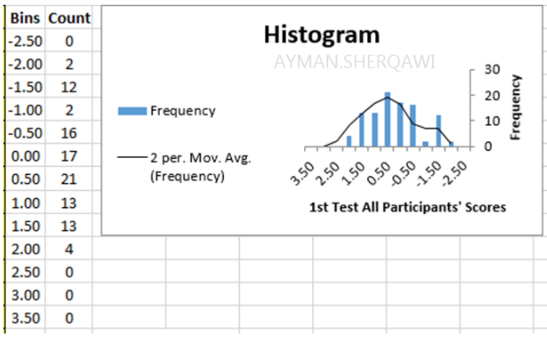
Bins	Count
-2.50	0
-2.00	0
-1.50	2
-1.00	0
-0.50	12
0.00	0
0.50	15
1.00	0
1.50	9
2.00	0
2.50	1
3.00	0
3.50	1

Figure 49 - Normal Distribution test for 3rd Test scores of the Online Group using Histograms and Bins



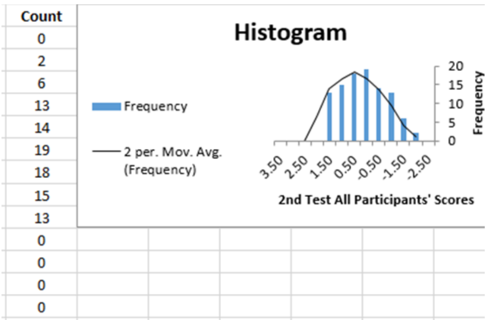
Count
0
0
9
0
20
0
21
0
7
0
2
0
1

Figure 50 - Normal Distribution test for 3rd Test scores of the Face-to-Face Group using Histograms and Bins



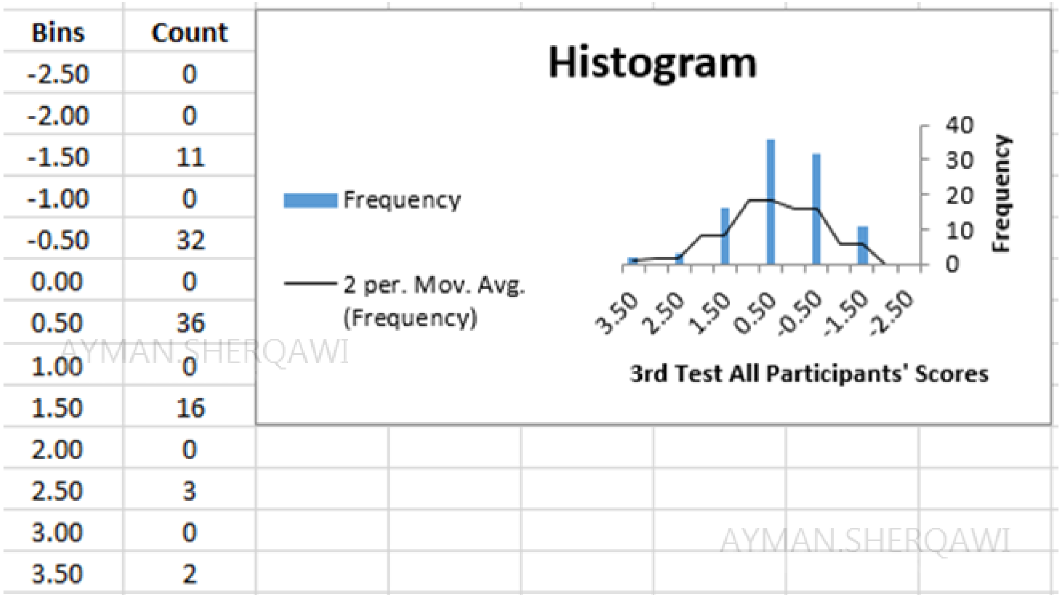
Bins	Count
-2.50	0
-2.00	2
-1.50	12
-1.00	2
-0.50	16
0.00	17
0.50	21
1.00	13
1.50	13
2.00	4
2.50	0
3.00	0
3.50	0

Figure 51 - Normal Distribution test for 1st Test scores of all participants combined using Histograms and Bins



Count
0
2
6
13
14
19
18
15
13
0
0
0
0

Figure 52 - Normal Distribution test for 2nd Test scores of all participants combined using Histograms and Bins



Bins	Count
-2.50	0
-2.00	0
-1.50	11
-1.00	0
-0.50	32
0.00	0
0.50	36
1.00	0
1.50	16
2.00	0
2.50	3
3.00	0
3.50	2

Figure 53 - Normal Distribution test for 3rd Test scores of all participants combined using Histograms and Bins

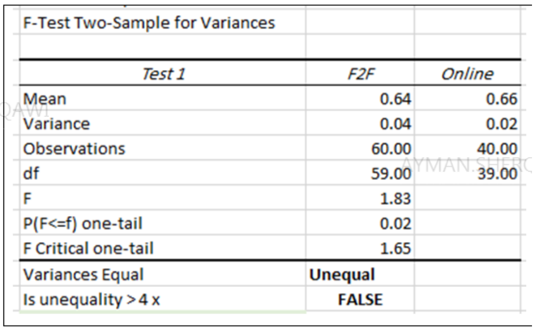
As illustrated by the frequency bins and histograms above, it is concluded that all tests' scores for both groups of participants have relatively close degrees of normal distributions. However, the histograms of the Online Participant group indicate that the Online Group consisting of 40 participants' scores tend to have a higher degree of normal distribution than that of the Face-to-Face Group which consists of 60 participants.

A.4 Variance Equality Testing

Another attribute of the data which is required by some of the parametric statistical tests namely the Independent t-test and Analysis of Variance (ANOVA) test is variance equality between the mean scores of the compared tests (Ruxton, 2006; Febrero and Fraiman, 2004). There are two types of the Independent t-test that deal with data with equal variances and data with unequal variances respectively. Thus, testing the variances of the compared independent samples for equality is a requirement to determine the right type of Intendent t-test to use. In order to test variance equality between any two independent datasets, an F-test (Bernhardson, 1975) is used. An F-test is used to determine whether a pair of independent samples (Face-to-Face Group dataset versus Online Group dataset) which are compared for differences in mean scores, have equal variances (Box, 1953; Levene, 1960). As such, three F-tests were conducted to compare the variances between the weighted mean scores of the Face-to-Face participant group dataset versus the Online participant group dataset for the 1st, 2nd and 3rd Tests respectively to determine if the variances between each pair of compared datasets were equal and hence eligible to use the Independent t-test and ANOVA test assuming equal variances.

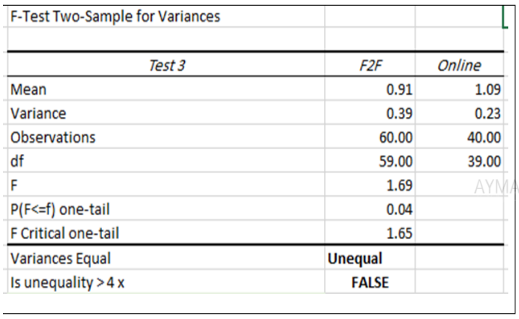
Thus, running F-test to check for variance equality between the weighted mean scores of the Online Participant group and the Face-to-Face Participant group on the 1st Test (Face-to-Face variance = 0.04, Online variance = 0.02) indicates unequal variances since (F = 1.83 > F Critical Value = 1.65). Similarly, running F-test to check the variances between the two groups of participants on the 3rd Test (Face-to-Face variance = 0.39, Online variance = 0.23) also indicates unequal variances since (F = 1.69 > F Critical Value = 1.65). However running F-test on the 2nd Test (Face-to-Face variance = 0.30, Online variance = 0.29) indicates equal variances since (F = 1.01 < F Critical Value = 1.65). On the 2nd Test, the F Statistic of (1.01) is less than the F Critical Value of (1.65) indicating equal variances between the Face-to-Face Group's weighted mean score (1.10) and the weighted mean score of the Online Group's (1.30) respectively (Markowski and Markowski, 1990) as shown in the generated F-test results report in Tables 50 through 52.

Table 50 - F-test results for testing variance equality between the Face-to-Face and Online groups on 1st Test



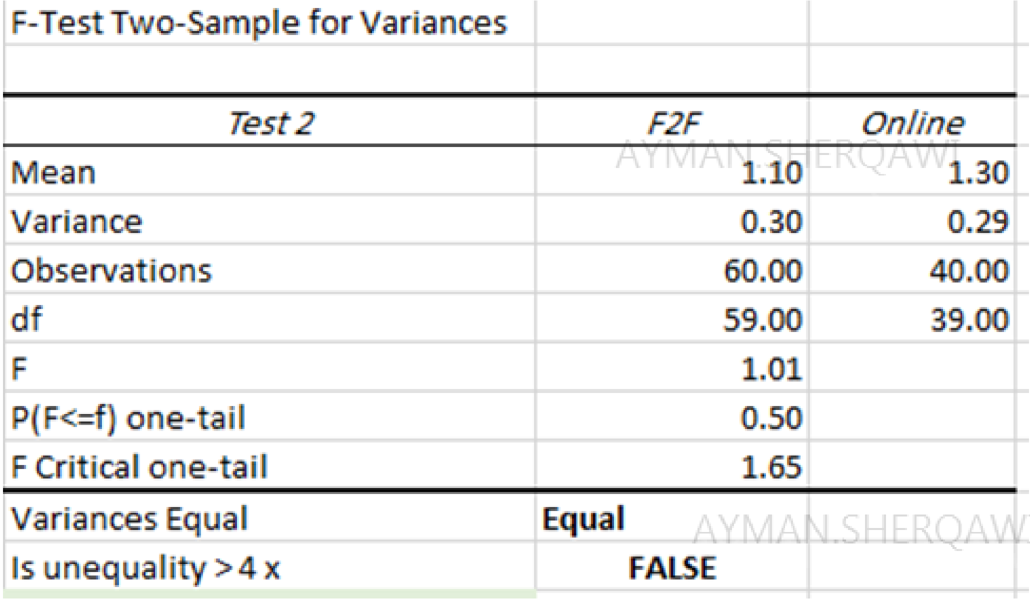
F-Test Two-Sample for Variances		
Test 1	*F2F*	*Online*
Mean	0.64	0.66
Variance	0.04	0.02
Observations	60.00	40.00
df	59.00	39.00
F	1.83	
P(F<=f) one-tail	0.02	
F Critical one-tail	1.65	
Variances Equal	**Unequal**	
Is unequality > 4 x	**FALSE**	

Table 51 - F-test results for testing variance equality between the Face-to-Face and Online groups on 3rd Test



F-Test Two-Sample for Variances		
Test 3	*F2F*	*Online*
Mean	0.91	1.09
Variance	0.39	0.23
Observations	60.00	40.00
df	59.00	39.00
F	1.69	
P(F<=f) one-tail	0.04	
F Critical one-tail	1.65	
Variances Equal	**Unequal**	
Is unequality > 4 x	**FALSE**	

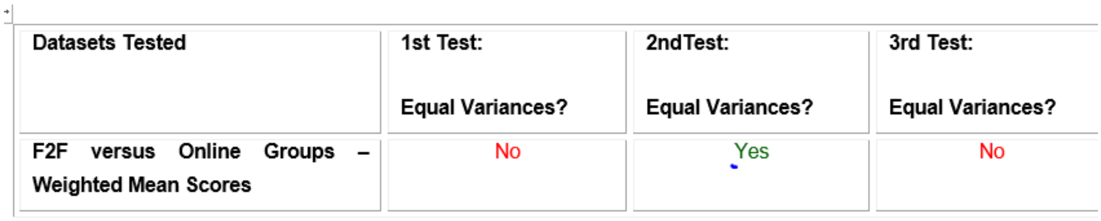
Table 52 - F-test results for testing variance equality between the Face-to-Face and Online groups on 2nd Test



F-Test Two-Sample for Variances		
Test 2	*F2F*	*Online*
Mean	1.10	1.30
Variance	0.30	0.29
Observations	60.00	40.00
df	59.00	39.00
F	1.01	
P(F<=f) one-tail	0.50	
F Critical one-tail	1.65	
Variances Equal	**Equal**	
Is unequality > 4 x	**FALSE**	

Despite the negative results of the F-tests on the 1st and 3rd Tests above indicating unequal variances, a less strict variance equality test was also conducted which yielded positive results indicating that the variances on the 1st and 3rd Tests are still within an acceptable range of equality (O'brien, 2007). This test checks to verify whether the variance of the first group (Face-to-Face Group) is not greater than (<) 4 times the second group's variance (Online Group) * (4). Hence, this rule holds true for the 1st and 3rd Tests. In conclusion, the results of running the F-tests as summarized in Table 53 indicate that the compared independent samples' datasets have equal variances only on the 2nd Test which qualifies them to be consumed by the Independent t-Test assuming Equal Variances. While on the 1st and 3rd Tests, unequal variances resulted indicating the feasibility of using the other type of Independent t-Test assuming Unequal variances.

Table 53 - Summary of F-Test results for variance equality for all three tests

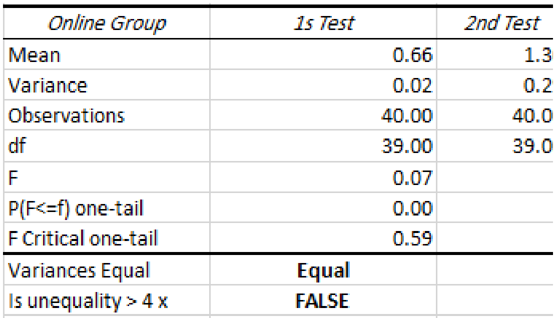


Datasets Tested	1st Test: Equal Variances?	2ndTest: Equal Variances?	3rd Test: Equal Variances?
F2F versus Online Groups – Weighted Mean Scores	No	Yes	No

Variance Equality for ANOVA: In addition to the checks above with regards to using the Independent t-test, using ANOVA test to compare pairs of datasets for a specific group of participants' scores between the three tests also requires equal variances. An example of this is the comparison of the Online Group's scores between the three Tests using ANOVA to identify if there is a significant difference between the three sets of scores. Therefore, 9 additional F-tests were conducted to compare the variances of each participant group's scores between the following paired tests: the 1st Test and 2nd Test, the 2nd Test and the 3rd Test and finally between the 3rd Test and 1st Test.

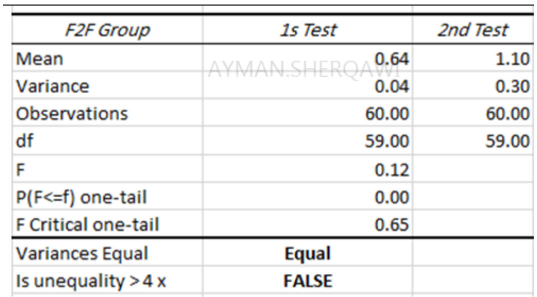
Thus, running F-test to check for variance equality between the 1st and 2nd Tests' weighted mean scores of the Online Participant group (1st Test Online Group variance = 0.02, 2nd Test Online Group variance = 0.29) indicates equal variances since (F = 0.07 < F Critical Value = 0.59). Similarly, running the same F-test on the Face-to-Face Participants' weighted scores (1st Test Face-to-Face Group variance = 0.04, 2nd Test Face-to-Face Group variance = 0.30) also indicates equal variances since (F = 0.12 < F Critical Value = 0.65). Equally, running F-Test on all participants combining both groups (1st Test All Participants variance = 0.03, 2nd Test All Participants variance = 0.30) still indicates equal variances since (F = 0.10 < F Critical Value = 0.72). Tables 54, 55 and 56 show the F-Tests results accordingly.

Table 54- F-test results for variance equality between 1st and 2nd Tests scores for the Online Group



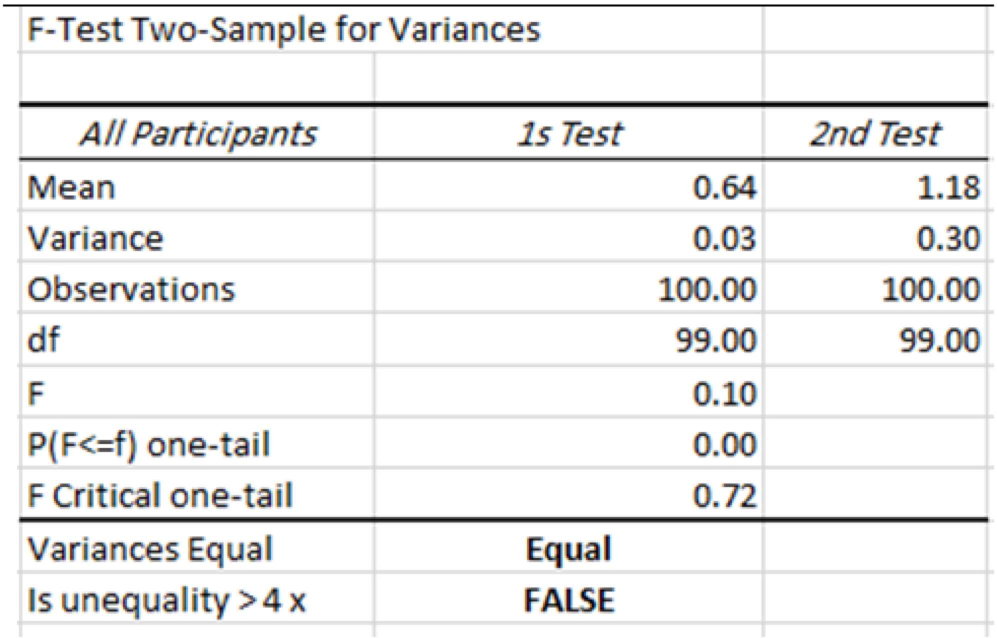
Online Group	*1s Test*	*2nd Test*
Mean	0.66	1.3
Variance	0.02	0.2
Observations	40.00	40.0
df	39.00	39.0
F	0.07	
P(F<=f) one-tail	0.00	
F Critical one-tail	0.59	
Variances Equal	**Equal**	
Is unequality > 4 x	**FALSE**	

Table 55 - F-test results for variance equality between 1st and 2nd Tests scores for the Face-to-Face Group



F2F Group	*1s Test*	*2nd Test*
Mean	0.64	1.10
Variance	0.04	0.30
Observations	60.00	60.00
df	59.00	59.00
F	0.12	
P(F<=f) one-tail	0.00	
F Critical one-tail	0.65	
Variances Equal	**Equal**	
Is unequality > 4 x	**FALSE**	

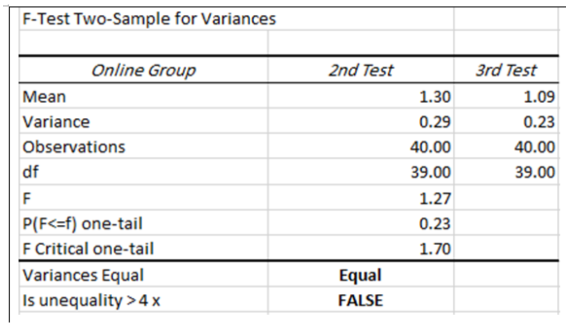
Table 56 - F-test results for variance equality between 1st and 2nd Tests scores for all participants combined



F-Test Two-Sample for Variances		
All Participants	*1s Test*	*2nd Test*
Mean	0.64	1.18
Variance	0.03	0.30
Observations	100.00	100.00
df	99.00	99.00
F	0.10	
P(F<=f) one-tail	0.00	
F Critical one-tail	0.72	
Variances Equal	**Equal**	
Is unequality > 4 x	**FALSE**	

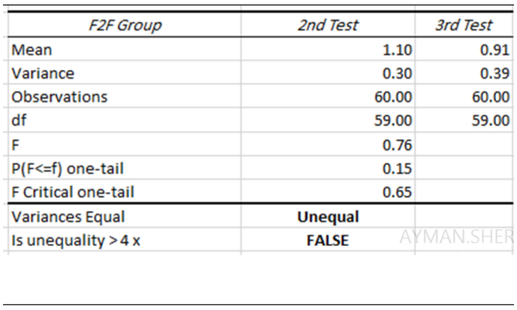
In the same way, running F-test to check for variance equality between the 2nd and 3rd Tests' weighted mean scores of the Online Participant group (2nd Test Online Group variance = 0.29, 3rd Test Online Group variance = 0.23) indicates equal variances since (F = 1.27 < F Critical Value = 1.70). However, running the same F-test on the Face-to-Face Participants' weighted scores (2nd Test Face-to-Face Group variance = 0.30, 3rd Test Face-to-Face Group variance = 0.39) indicates unequal variances since (F = 0.76 > F Critical Value = 0.65). Overall, running F-Test on all participants combining both groups (2nd Test All Participants variance = 0.30, 3rd Test All Participants variance = 0.33) yet indicates unequal variances since (F = 0.91 > F Critical Value = 0.72). Tables 57, 58 and 59 show the F-Tests results accordingly. Despite the results of the F-tests on the Face-to-Face participants in particular and the overall participants on the 2nd and 3rd Tests indicating unequal variances, these variances are still within an acceptable range of equality since the variances on the 2nd Test are not greater than 4 times the variances on the 3rd Test (O'brien, 2007) for the Face-to-Face Group and for all participants in general. Therefore, these two groups of participants' scores are considered to have equal variances between the 2nd and 3rd Tests.

Table 57 - F-test results for variance equality between 2nd and 3rd Tests scores for the Online Group



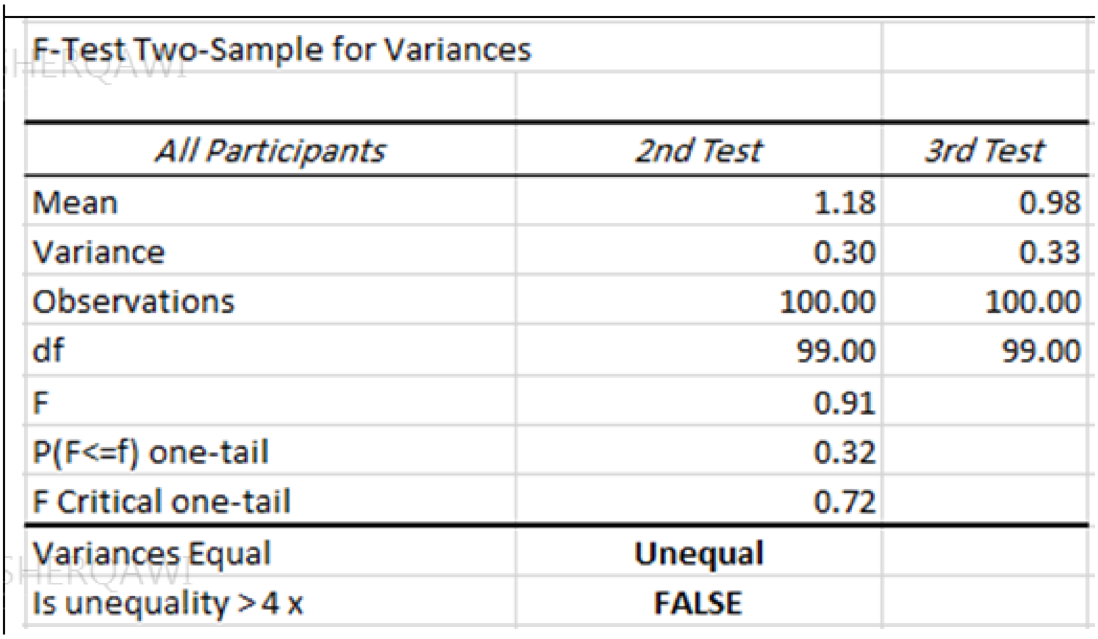
F-Test Two-Sample for Variances		
Online Group	*2nd Test*	*3rd Test*
Mean	1.30	1.09
Variance	0.29	0.23
Observations	40.00	40.00
df	39.00	39.00
F	1.27	
P(F<=f) one-tail	0.23	
F Critical one-tail	1.70	
Variances Equal	**Equal**	
Is unequality > 4 x	**FALSE**	

Table 58 - F-test results for variance equality between 2nd and 3rd Tests scores for the Face-to-Face Group



F2F Group	*2nd Test*	*3rd Test*
Mean	1.10	0.91
Variance	0.30	0.39
Observations	60.00	60.00
df	59.00	59.00
F	0.76	
P(F<=f) one-tail	0.15	
F Critical one-tail	0.65	
Variances Equal	**Unequal**	
Is unequality > 4 x	**FALSE**	

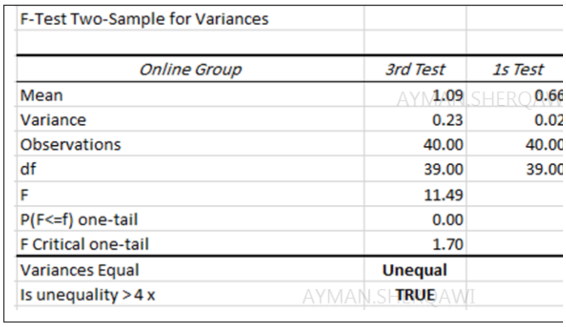
Table 59 - F-test results for variance equality between 2nd and 3rd Tests scores for all participants combined



F-Test Two-Sample for Variances		
All Participants	*2nd Test*	*3rd Test*
Mean	1.18	0.98
Variance	0.30	0.33
Observations	100.00	100.00
df	99.00	99.00
F	0.91	
P(F<=f) one-tail	0.32	
F Critical one-tail	0.72	
Variances Equal	**Unequal**	
Is unequality > 4 x	**FALSE**	

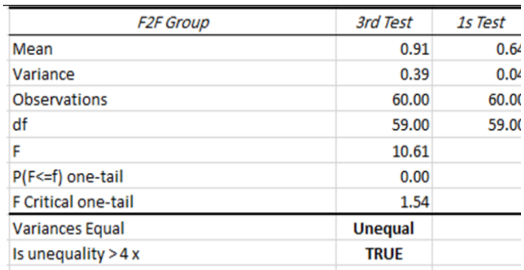
Finally, running F-test to check for variance equality between the 3rd and 1st Tests' weighted mean scores of the Online Participant group (3rd Test Online Group variance = 0.23, 1st Test Online Group variance = 0.02) indicates unequal variances since (F = 11.49 > F Critical Value = 1.70). Similarly, running the F-test on the Face-to-Face Participants' weighted scores (3rd Test Face-to-Face Group variance = 0.39, 1st Test Face-to-Face Group variance = 0.04) also indicates unequal variances since (F = 10.61 > F Critical Value = 1.54). Overall, running F-Test on all participants combining both groups (3rd Test All Participants variance = 0.33, 1st Test All Participants variance = 0.03) still indicates unequal variances since (F = 11.06 > F Critical Value = 1.39). Tables 60, 61 and 62 show the F-Tests results accordingly.

Table 60 - F-test results for variance equality between 3rd and 1st Tests scores for the Online Group



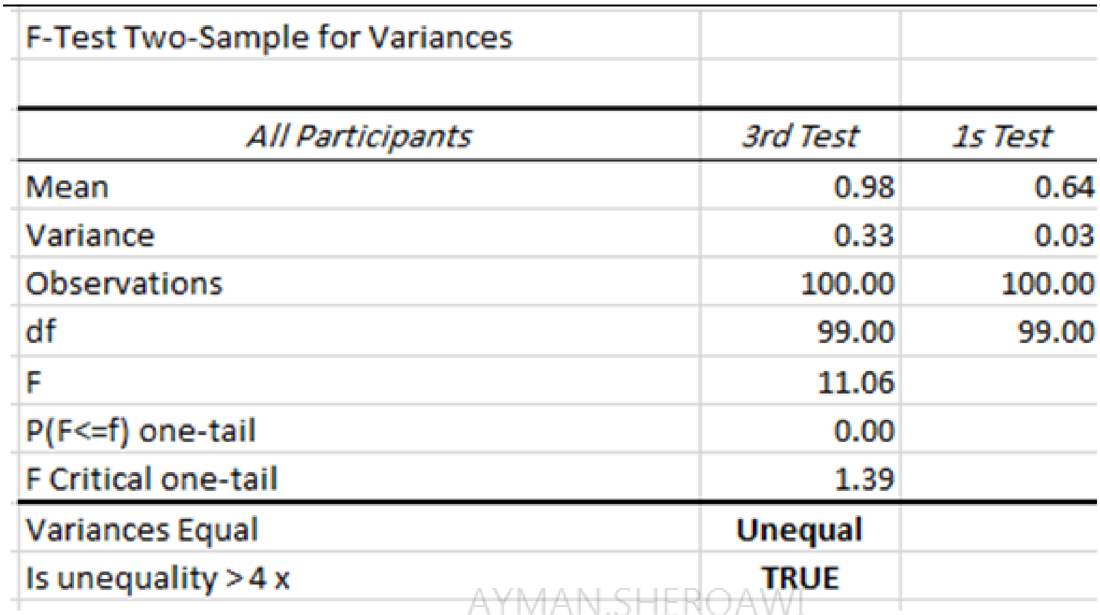
F-Test Two-Sample for Variances		
Online Group	*3rd Test*	*1s Test*
Mean	1.09	0.66
Variance	0.23	0.02
Observations	40.00	40.00
df	39.00	39.00
F	11.49	
P(F<=f) one-tail	0.00	
F Critical one-tail	1.70	
Variances Equal	**Unequal**	
Is unequality > 4 x	**TRUE**	

Table 61 - F-test results for variance equality between 3rd and 1st Tests scores for the Face-to-Face Group



F2F Group	*3rd Test*	*1s Test*
Mean	0.91	0.64
Variance	0.39	0.04
Observations	60.00	60.00
df	59.00	59.00
F	10.61	
P(F<=f) one-tail	0.00	
F Critical one-tail	1.54	
Variances Equal	**Unequal**	
Is unequality > 4 x	**TRUE**	

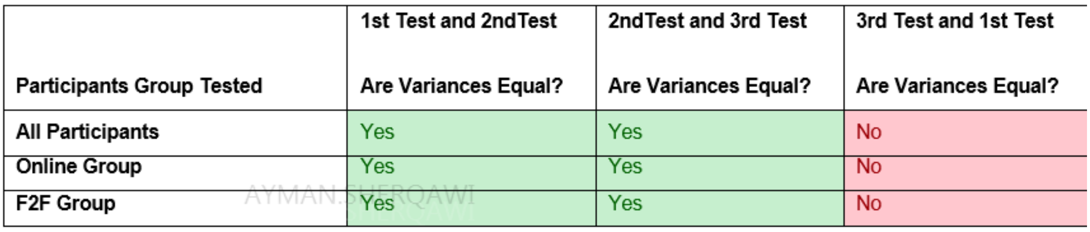
Table 62 - F-test results for variance equality between 3rd and 1st Tests scores for all participants combined



F-Test Two-Sample for Variances		
All Participants	*3rd Test*	*1s Test*
Mean	0.98	0.64
Variance	0.33	0.03
Observations	100.00	100.00
df	99.00	99.00
F	11.06	
P(F<=f) one-tail	0.00	
F Critical one-tail	1.39	
Variances Equal	**Unequal**	
Is unequality > 4 x	**TRUE**	

In conclusion, the results of the F-tests conducted above summarized in Table 63 indicate that the variances between the 1st Test and 2nd Test and between the 2nd Test and 3rd Test for all the participant groups are equal and that the assumption of variance equality is met and hence qualifies the datasets of these groups to be consumed by the ANOVA test. However, the variances between the scores of the 3rd Test and the 1st Test are unequal rendering ANOVA test inappropriate for use with these datasets. In that case, the non-parametric test kruskal-Wallis H test (Wallace, 1995) will be used as an alternative test.

Table 63 - Summary of all F-tests' results



Participants Group Tested	1st Test and 2ndTest Are Variances Equal?	2ndTest and 3rd Test Are Variances Equal?	3rd Test and 1st Test Are Variances Equal?
All Participants	Yes	Yes	No
Online Group	Yes	Yes	No
F2F Group	Yes	Yes	No

Now that the data have been checked for normality and variance equality, the appropriate statistical tests are determined to start the hypotheses testing in Chapter 5 Sections 11-17.

Appendix B – List of the 18 Hypotheses

Null Hypothesis (1a): There is no significant difference between the mean scores of the Online Group and the Face-to-Face Group on the 1st Test. In other words, there is no significant difference between the Online participants and the Face-to-Face participants in the 'Perception' level of awareness about Phishing.

> **Alternate Hypothesis (1b)**: There is a significant difference between the mean scores of the Online Group and the Face-to-Face Group on the 1st Test. In other words, there is a statistically significant difference between the Online participants and the Face-to-Face participants in the 'Perception' level of awareness about Phishing as statistically manifested by the 1st Test's scores. If the result of the statistical test does not significantly support the Null hypothesis and hence supports the Alternate hypothesis, the participant group with the higher mean score will be the group whose participants have a better Perception awareness level than that of the other group.

Null Hypothesis (2a): There is no significant difference between the mean scores of the Online Group and the Face-to-Face Group on the 2nd Test. In other words, there is no significant difference between the Online participants and the Face-to-Face participants in the 'Comprehension' level of awareness about Phishing after attending the 1st awareness session.

> **Alternate Hypothesis (2b):** There is a significant difference between the mean scores of the Online Group and the Face-to-Face Group on the 2nd Test. In other words, there is a statistically significant difference between the Online participants and the Face-to-Face participants in the 'Comprehension' level of awareness about Phishing after attending the 1st awareness session.

Null Hypothesis (3a): There is no significant difference between the mean scores of the Online Group and the Face-to-Face Group on the 3rd Test. In other words, there is no significant difference between the Online participants and the Face-to-Face participants in the 'Projection' level of awareness about Phishing after attending the 2nd awareness session.

Alternate Hypothesis (3b): There is a significant difference between the mean scores of the Online Group and the Face-to-Face Group on the 3rd Test. In other words, there is a statistically significant difference between the Online participants and the Face-to-Face participants in the 'Projection' level of awareness about Phishing after attending the 2nd awareness session.

Null Hypothesis (4a): There is no significant difference between the mean scores acquired by the Online participant group on the 1st, 2nd and 3rd Tests. In other words, there is no statistically significant difference between the three tests' scores for the Online participant group that will indicate gradual improvements in participant awareness from the Perception Level to the Comprehension level and then finally to the Projection Level after attending the two awareness sessions respectively.

Alternate Hypothesis (4b): There is a significant difference between the mean scores acquired by the Online participant group on the 1st, 2nd and 3rd Tests. In other words, there is a statistically significant difference between the three tests' scores for the Online participant group that will indicate gradual improvements in participant awareness from the Perception Level to the Comprehension level and then finally to the Projection Level after attending the two awareness sessions respectively.

Null Hypothesis (5a): There is no significant difference between the mean scores acquired by the Face-to-Face participant group on the 1st, 2nd and 3rd Tests. In other words, there is no statistically significant difference between the three tests' scores for the Face-to-Face participant group that will indicate gradual improvements in participant awareness from the Perception Level to the Comprehension level and then finally to the Projection Level after attending the two awareness sessions respectively.

Alternate Hypothesis (5b): There is a significant difference between the mean scores acquired by the Face-to-Face participant group on the 1st, 2nd and 3rd Tests. In other words, there is a statistically significant difference between the three tests' scores for the Face-to-Face participant group that will indicate gradual improvement in participant awareness from the Perception Level to the Comprehension level and then finally to the Projection Level after attending the two awareness sessions respectively.

Null Hypothesis (6a): There is no significant difference between the mean scores acquired by all participants on the 1st, 2nd and 3rd Tests. In other words, there is no statistically significant difference between the three tests' scores for all participants that will indicate gradual improvements in participant awareness from the Perception Level to the Comprehension level and then finally to the Projection Level after attending the two awareness sessions respectively.

Alternate Hypothesis (6b): There is a significant difference between the mean scores acquired by all participants on the 1st, 2nd and 3rd Tests. In other words, there is a statistically significant difference between the three tests' scores for all participants that will indicate gradual improvements in participant awareness from the Perception Level to the Comprehension level and then finally to the Projection Level after attending the two awareness sessions respectively.

Null Hypothesis (7a): There is no statistically significant difference between the mean scores on the 2nd Test and the 1st Test for all participants. In other words, the mean scores on the 2nd Test is not significantly higher or lower than the mean scores on the 1st Test for all participants.

Alternate Hypothesis (7b): There is a statistically significant difference between the mean scores on the 2nd Test and the 1st Test for all participants. In other words, the mean scores on the 2nd Test is significantly higher or lower than the mean scores on the 1st Test for all participants.

Null Hypothesis (8a): There is no statistically significant difference between the mean scores on the 2nd Test and the 1st Test for the Online participant group. In other words, the mean scores on the 2nd Test is not significantly higher or lower than the mean scores on the 1st Test for the Online participant group.

> **Alternate Hypothesis (8b):** There is a statistically significant difference between the mean scores on the 2nd Test and the 1st Test for the Online participant group. In other words, the mean scores on the 2nd Test is significantly higher or lower than the mean scores on the 1st Test for the Online participant group.

Null Hypothesis (9a): There is no statistically significant difference between the mean scores on the 2nd Test and the 1st Test for the Face-to-Face participant group. In other words, the mean scores on the 2nd Test is not significantly higher or lower than the mean scores on the 1st Test for the Face-to-Face participant group.

> **Alternate Hypothesis (9b):** There is a statistically significant difference between the mean scores on the 2nd Test and the 1st Test for the Face-to-Face participant group. In other words, the mean scores on the 2nd Test is significantly higher or lower than the mean scores on the 1st Test for the Face-to-Face participant group.

Null Hypothesis (10a): There is no statistically significant difference between the mean scores on the 3rd Test and the 2nd Test for all participants. In other words, the mean scores on the 3rd Test is not significantly higher or lower than the mean scores on the 2nd Test for all participants.

> **Alternate Hypothesis (10b)**: There is a statistically significant difference between the mean scores on the 3rd Test and the 2nd Test for all participants. In other words, the mean scores on the 3rd Test is significantly higher or lower than the mean scores on the 2nd Test for all participants.

Null Hypothesis (11a): There is no statistically significant difference between the mean scores on the 3rd Test and the 2nd Test for the Online participant group. In other words, the mean score on the 3rd Test is not significantly higher or lower than the mean scores on the 2nd Test for the Online participant group.

> **Alternate Hypothesis (11b):** There is a statistically significant difference between the mean scores on the 3rd Test and the 2nd Test for the Online participant group. In other words, the mean scores on the 3rd Test is significantly higher or lower than the mean scores on the 2nd Test for the Online participant group.

Null Hypothesis (12a): There is no statistically significant difference between the mean scores on the 3rd Test and the 2nd Test for the Face-to-Face participant group. In other words, the mean scores on the 3rd Test is not significantly higher or lower than the mean scores on the 2nd Test for the Face-to-Face participant group.

> **Alternate Hypothesis (12b):** There is a statistically significant difference between the mean scores on the 3rd Test and the 2nd Test for the Face-to-Face participant group. In other words, the mean scores on the 3rd Test is significantly higher or lower than the mean scores on the 2nd Test for the Face-to-Face participant group.

Null Hypothesis (13a): There is no statistically significant difference between the mean scores on the 3rd Test and the 1st Test for all participants. In other words, the mean scores on the 3rd Test is not significantly higher or lower than the mean scores on the 1st Test for all participants.

> **Alternate Hypothesis (13b):** There is a statistically significant difference between the mean scores on the 3rd Test and the 1st Test for all participants. In other words, the mean scores on the 3rd Test is significantly higher or lower than the mean scores on the 1st Test for all participants.

Null Hypothesis (14a): There is no statistically significant difference between the mean scores on the 3rd Test and the 1st Test for the Online participant group. In other words, the mean scores on the 3rd Test is not significantly higher or lower than the mean scores on the 1st Test for the Online participant group.

> **Alternate Hypothesis (14b):** There is a statistically significant difference between the mean scores on the 3rd Test and the 1st Test for the Online participant group. In other words, the mean scores on the 3rd Test is significantly higher or lower than the mean scores on the 1st Test for the Online participant group.

Null Hypothesis (15a): There is no statistically significant difference between the mean scores on the 3rd Test and the 1st Test for the Face-to-Face participant group. In other words, the mean scores on the 3rd Test is not significantly higher or lower than the mean scores on the 1st Test for the Face-to-Face participant group.

> **Alternate Hypothesis (15b):** There is a statistically significant difference between the mean scores on the 3rd Test and the 1st Test for the Face-to-Face participant group. In other words, the mean scores on the 3rd Test is significantly higher or lower than the mean scores on the 1st Test for the Face-to-Face participant group.

Null Hypothesis (16a): The independent variables (Participants' Age, Education level and Years of online study) do not significantly impact the scores (dependent variable) on the 1st Test. In other words, participants' attributes do not significantly affect the scores on the 1st Test and hence cannot accurately be used to predict participants' scores on future tests.

> **Alternate Hypothesis (16b):** The independent variables (Participants' Age, Education level and Years of online study) significantly impact the scores (dependent variable) on the 1st Test. In other words, participants' attributes can significantly affect the scores on the 1st Test and hence can be used to predict participants' scores on future tests.

Null Hypothesis (17a): The independent variables (Participants' Age, Education level and Years of online study) do not significantly impact the scores (dependent variable) on the 2nd Test. In other words, participants' attributes do not significantly affect the scores on the 2nd Test and hence cannot accurately be used to predict participants' scores on future tests.

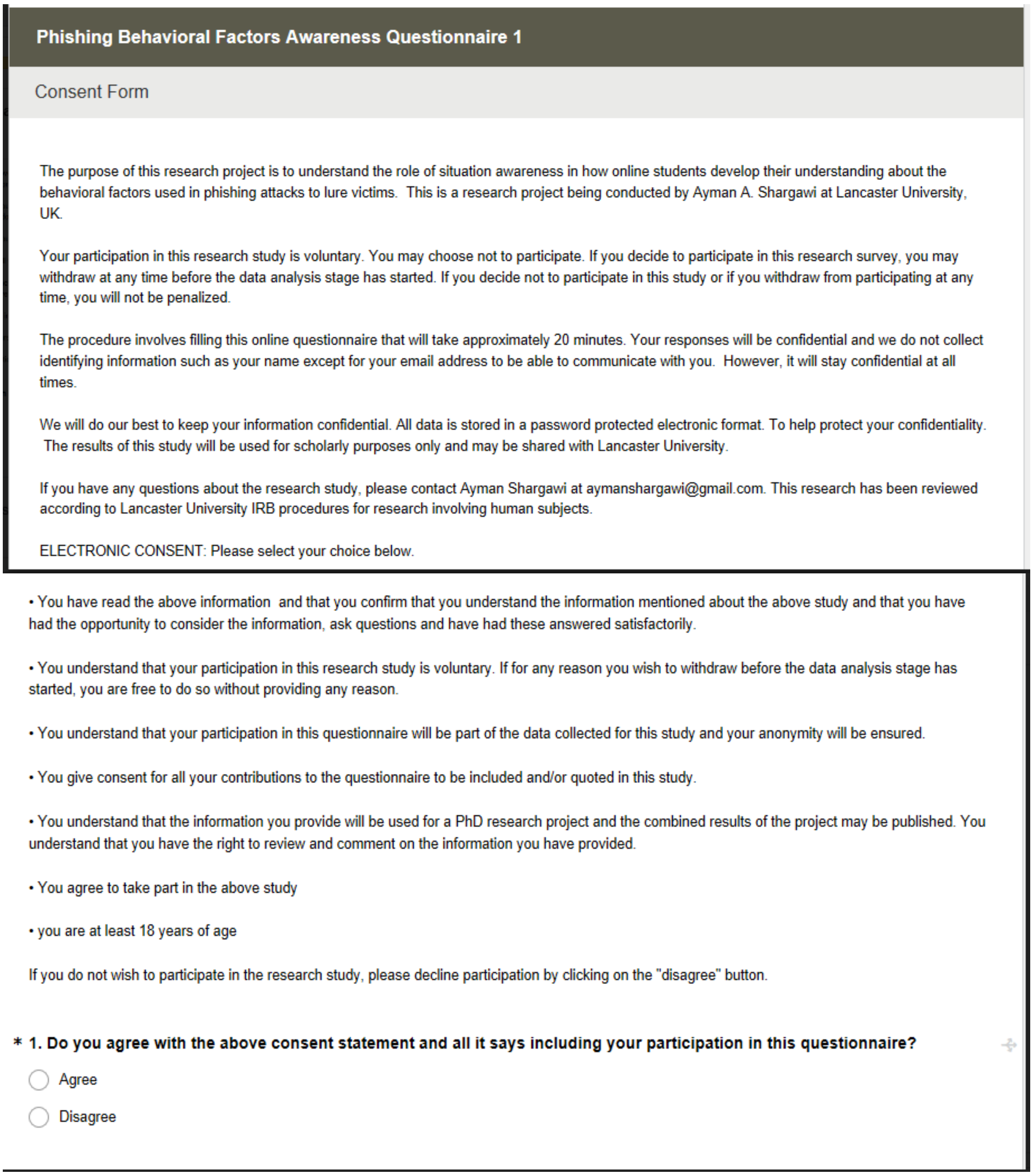
> **Alternate Hypothesis (17b):** The independent variables (Participants' Age, Education level and Years of online study) significantly impact the scores (dependent variable) on the 2nd Test. In other words, participants' attributes can significantly affect the scores on the 2nd Test and hence can be used to predict participants' scores on future tests.

Null Hypothesis (18a): The independent variables (Participants' Age, Education level and Years of online study) do not significantly impact the scores (dependent variable) on the 3rd Test. In other words, participants' attributes do not significantly affect the scores on the 3rd Test and hence cannot accurately be used to predict participants' scores on future tests.

> **Alternate Hypothesis (18b):** The independent variables (Participants' Age, Education level and Years of online study) significantly impact the scores (dependent variable) on the 3rd Test. In other words, participants' attributes can significantly affect the scores on the 3rd Test and hence can be used to predict participants' scores on future tests.

Appendix C – Consent Form

Below is a screenshot of the online consent form participants signed before starting their participation in this research. This consent form is part of the 1st Test that was developed using SurveyMonkey.



Phishing Behavioral Factors Awareness Questionnaire 1

Consent Form

The purpose of this research project is to understand the role of situation awareness in how online students develop their understanding about the behavioral factors used in phishing attacks to lure victims. This is a research project being conducted by Ayman A. Shargawi at Lancaster University, UK.

Your participation in this research study is voluntary. You may choose not to participate. If you decide to participate in this research survey, you may withdraw at any time before the data analysis stage has started. If you decide not to participate in this study or if you withdraw from participating at any time, you will not be penalized.

The procedure involves filling this online questionnaire that will take approximately 20 minutes. Your responses will be confidential and we do not collect identifying information such as your name except for your email address to be able to communicate with you. However, it will stay confidential at all times.

We will do our best to keep your information confidential. All data is stored in a password protected electronic format. To help protect your confidentiality. The results of this study will be used for scholarly purposes only and may be shared with Lancaster University.

If you have any questions about the research study, please contact Ayman Shargawi at aymanshargawi@gmail.com. This research has been reviewed according to Lancaster University IRB procedures for research involving human subjects.

ELECTRONIC CONSENT: Please select your choice below.

• You have read the above information and that you confirm that you understand the information mentioned about the above study and that you have had the opportunity to consider the information, ask questions and have had these answered satisfactorily.

• You understand that your participation in this research study is voluntary. If for any reason you wish to withdraw before the data analysis stage has started, you are free to do so without providing any reason.

• You understand that your participation in this questionnaire will be part of the data collected for this study and your anonymity will be ensured.

• You give consent for all your contributions to the questionnaire to be included and/or quoted in this study.

• You understand that the information you provide will be used for a PhD research project and the combined results of the project may be published. You understand that you have the right to review and comment on the information you have provided.

• You agree to take part in the above study

• you are at least 18 years of age

If you do not wish to participate in the research study, please decline participation by clicking on the "disagree" button.

*** 1. Do you agree with the above consent statement and all it says including your participation in this questionnaire?**

○ Agree

○ Disagree

References

Abramowitz, M., and Stegun, I. A. (1967). *Formulas*. Graphs, and Mathematical Tables.

Adams, J. E. (2012). Mutual authentication & phishing – a clear case of user confusion. Retrieved from http://www.jane-adams.com/wp-content/uploads/2012/09/janeadamsdissertation.pdf

Aiken, L. R. (1987). Formulas for equating ratings on different scales. Educational and Psychological Measurement, 47(1), 51–54. doi:10.1177/0013164487471007

Alecu, F., Pocatilu, P., & Capisizu, S. (2010). WiMAX Security issues in e-Learning systems. Journal of Mobile. Embedded and Distributed Systems, 2(1), 15–20.

Aljawarneh, S. (2011). A web engineering security methodology for e-learning systems. Network Security, 2011(3), 12–15. doi:/10.1016/S1353-4858(11)70026-5

Allen, I. E., & Seaman, C. A. (2007). Likert scales and data analyses. Quality progress, 40(7), 64.

Altman, D. G., & Bland, J. M. (1996). Detecting skewness from summary information. BMJ (Clinical Research Ed.), 313(7066), 1200. doi.org/10.1136/bmj.313.7066.1200

Anandpara, V., Dingman, A., Jakobsson, M., Liu, D., & Roinestad, H. (2007). Phishing IQ tests measure fear, not ability. Lecture Notes in Computer Science, 4886, 362-366.

Andriessen, D. (2006, August). Combining design-based research and action research to test management solutions. In 7th World Congress Action Research.

Anttila, J., Savola, R., Kajava, J., Lindfors, J., & Röning, J. (2007). Fulfilling the needs for information security awareness and learning in information society. In The 6th Annual Security Conference, Las Vegas.

Bandyopadhyay, S. Normal Probability plot. AMS.

Bansode, U. M., & Rao, G. R. (2013). Study of Various Anti-Phishing Approaches and Introducing an Approved Method for Detecting Phishing Websites. International Journal on Computer Science and Engineering, 1(2), 151–156.

Barik, N., & Karforma, S. (2012). Risks and remedies in e-learning system.arXiv preprint arXiv: 1205.2711.

Bernhardson, C. S. (1975). 375: Type I error rates when multiple comparison procedures follow a significant F test of ANOVA. Biometrics, 31(1), 229–232. doi.org/10.2307/2529724

Bai, J., & Ng, S. (2005). Tests for skewness, kurtosis, and normality for time series data. Journal of Business and Economic Statistics, 23(1), 49–60. doi.org/10.1198/073500104000000271

Box, G. E. (1953). Non-normality and tests on variances. Biometrika, 40(3/4), 318–335. doi.org/10.2307/2333350

Braun, H. I., & Holland, P. W. (1982). Observed-score test equating: A mathematical analysis of some ETS equating procedures. Test equating, 9-49.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative Research in Psychology, 3(2), 77-101. doi:10.1191/1478088706qp063oa

Brehmer, B. (2005, June). The dynamic OODA loop: Amalgamating Boyd's OODA loop and the cybernetic approach to command and control. In Proceedings of the 10th International Command and Control Research Technology Symposium (pp. 365-368).

Brownlee, K. A., & Brownlee, K. A. (1965). Statistical theory and methodology in science and engineering (Vol. 150). New York: Wiley.

Brynielsson, J., Franke, U., & Varga, S. (2016). Cyber Situational Awareness Testing. In Combatting Cybercrime and Cyberterrorism (pp. 209-233). Springer International Publishing. doi.org/10.1007/978-3-319-38930-1_12

Burke, R. E., Fahn, S., Marsden, C. D., Bressman, S. B., Moskowitz, C., & Friedman, J. (1985). Validity and reliability of a rating scale for the primary torsion dystonias. Neurology, 35(1), 73–77. doi.org/10.1212/WNL.35.1.73

Caro, I. A. (2011, June). Vex Robotics: stem Program and Robotics Competition Expansion into Europe. In Proceedings of International Conference on Research and Education in Robotics (pp. 10-16). Berlin Heidelberg: Springer.

Chen, C. C., Shaw, R. S., & Yang, S. C. (2006). Mitigating information security risks by increasing user security awareness: A case study of an information security awareness system. Information Technology, Learning and Performance Journal, 24(1), 1.

Chen, Y., & He, W. (2013). Security risks and protection in online learning: A survey. International Review of Research in Open and Distributed Learning, 14(5). doi.org/10.19173/irrodl.v14i5.1632

Chu, W., Zhu, B. B., Xue, F., Guan, X., & Cai, Z. (2013, June). Protect sensitive sites from phishing attacks using features extractable from inaccessible phishing URLs. In IEEE International Conference on Communications (pp. 1990-1994). IEEE. doi.org/10.1109/ICC.2013.6654816

Cohen, L., Manion, L., & Morrison, K. (2011). Research methods in education. London: RoutledgeFalmer.

Cohen, J. (1992). A power primer. Psychological Bulletin, 112(1), 155. Retrieved from

http://www2.psych.ubc.ca/~schaller/528Readings/Cohen1992.pdf

Cone, B. D., Irvine, C. E., Thompson, M. F., & Nguyen, T. D. (2007). A video game for cyber security training and awareness. Computers and security, 26(1), 63-72.

Conover, W. J., & Iman, R. L. (1981). Rank transformations as a bridge between parametric and nonparametric statistics. The American Statistician, 35(3), 124–129.

Costinela-Luminiţa, C. D., & Nicoleta-Magdalena, C. I. (2012). E-learning security vulnerabilities. Procedia: Social and Behavioral Sciences, 46, 2297–2301. doi.org/10.1016/j.sbspro.2012.05.474

Coyne, I. (1997). Sampling in qualitative research. Purposeful and theoretical sampling; merging or clear boundaries? Journal of Advanced Nursing, 26(3), 623-630. doi:10.1046/j.1365-2648.1997.t01-25-00999.x

Cuevas, A., Febrero, M., & Fraiman, R. (2004). An anova test for functional data. Computational Statistics and Data Analysis, 47(1), 111–122. doi.org/10.1016/j.csda.2003.10.021

Delice, A. (2010). The Sampling Issues in Quantitative Research. Educational Sciences: Theory and Practice, 10(4), 2001–2018.

Dixon, W. J., & Mood, A. M. (1946). The statistical sign test. Journal of the American Statistical Association, 41(236), 557–566. doi.org/10.1080/01621459.1946.10501898

Draper, N. R. (1966). Applied Regression Analysis. Retrieved from https://leseprobe.buch.de/images-adb/dd/cc/ddcc9caf-cd31-439b-8d9b-1080fd7fddc7.pdf

Elliot, J. (1991). Action research for educational change. UK: McGraw-Hill Education.

Ellis, T. J., & Levy, Y. (2008). Framework of problem-based research: A guide for novice researchers on the development of a research-worthy problem. Informing Science.

Endsley, M. R. (2000). Theoretical underpinnings of situation awareness: A critical review. Situation awareness analysis and measurement, 3-32.

Endsley, M. R. (2015). Situation awareness misconceptions and misunderstandings. Journal of Cognitive Engineering and Decision Making, 9(1), 4–32. doi.org/10.1177/1555343415572631

Etikan, I. (2016). Comparison of Convenience Sampling and Purposive Sampling. American Journal of Theoretical and Applied Statistics, 5(1), 1. doi:10.11648/j.ajtas.20160501.11

Ferreira, A., & Lenzini, G. (2015, July). An analysis of social engineering principles in effective phishing. In Workshop on Socio-Technical Aspects in Security and Trust, (pp. 9-16). IEEE. doi.org/10.1109/STAST.2015.10

Filliben, J. J. (1975). The probability plot correlation coefficient test for normality. Technometrics, 17(1), 111–117. doi.org/10.1080/00401706.1975.10489279

Fischer, R. (2004). Standardization to account for cross-cultural response bias a classification of score adjustment procedures and review of research in JCCP. Journal of Cross-Cultural Psychology, 35(3), 263–282. doi.org/10.1177/0022022104264122

Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (1993). How to design and evaluate research in education (Vol. 7). New York: McGraw-Hill.

Furnell, S. (2008). End-user security culture: A lesson that will never be learnt? Computer Fraud and Security, (4): 6–9.

Furnell, S. (2012). Disguising the dangers: Hiding attacks behind modern masks. Computer Fraud and Security, 2012(6), 9–13. doi.org/10.1016/S1361-3723(12)70062-3

Ghasemi, A., & Zahediasl, S. (2012). Normality tests for statistical analysis: A guide for non-statisticians. International Journal of Endocrinology and Metabolism, 10(2), 486–489. doi.org/10.5812/ijem.3505

Ghasemi, A., & Zahediasl, S. (2012). Normality tests for statistical analysis: A guide for non-statisticians. International Journal of Endocrinology and Metabolism, 10(2), 486–489. doi.org/10.5812/ijem.3505

Gibbons, J. D., & Chakraborti, S. (2011). Nonparametric statistical inference (pp. 977–979). Berlin Heidelberg: Springer.

Gillies, J., & Quijada, J. J. (2008). Opportunity to Learn: A High Impact Strategy for Improving Educational Outcomes in Developing Countries. Working Paper. Academy for Educational Development.

Goforth, C. (2015). Using and Interpreting Cronbach's Alpha. University of Virginia

Library. Retrieved from http://data.library.virginia.edu/using-and-interpreting-cronbachs-alpha/

Golafshani, N. (2003). Understanding reliability and validity in qualitative research. Qualitative Report, 8(4), 597–606.

Gunawardena, C. N. (1995). Social presence theory and implications for interaction and collaborative learning in computer conferences. International Journal of Educational Telecommunications, 1(2/3), 147–166.

Gupta, M., Sharman, Raj, & IGI Global. (2009). Social and human elements of information security [electronic resource]: Emerging trends and countermeasures. Hershey, Pa.: IGI Global.

Hanamura, K., Takemura, T., & Komatsu, A. (2013). Research Note: Analysis of the Characteristics of Victims in Information Security Incident Damages: The Case of Japanese Internet Users. The Review of Socionetwork Strategies, 7(1), 43–51. doi.org/10.1007/s12626-013-0032-6

Heale, R., & Twycross, A. (2015). Validity and reliability in quantitative studies. Evidence-Based Nursing, 18(3), 66–67. doi.org/10.1136/eb-2015-102129

Higbee, A., Belani, R., & Greaux, S. (2013). U.S. Patent No. 8,615,807. Washington, DC: U.S. Patent and Trademark Office.

Hong, K. W., Kelley, C. M., Tembe, R., Murphy-Hill, E., & Mayhorn, C. B. (2013). Keeping up with the Joneses: Assessing phishing susceptibility in an email task. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 57(1), 1012–1016. doi.org/10.1177/1541931213571226

Huang, H., Qian, L., & Wang, Y. (2012). A SVM-based technique to detect phishing URLs. Information Technology Journal, 11(7), 921–925. doi.org/10.3923/itj.2012.921.925

IBM, (2016, September 07). Retrieved from http://www-01.ibm.com/support/docview.wss?uid=swg21482329

Ibrahim, R. (2016). The Effect of Personality on SMS Phishing Vulnerability (Unpublished doctoral dissertation), University of York, UK).

Isaacson, K. (2013). An Investigation into the Affordances of Google Hangouts for possible use in Synchronous Online Learning Environments. In J. Herrington, A. Couros & V. Irvine (Eds.), Proceedings of World Conference on Educational Media and Technology 2013 (pp. 2461-2465). Association for the Advancement of Computing in Education.

Isaacson, K. (2013, June). An investigation into the affordances of google hangouts for possible use in synchronous online learning environments. In World Conference on.

Jadhav, A., & Sonar, R. (2009, December). Analytic Hierarchy Process (AHP), Weighted Scoring Method (WSM), and Hybrid Knowledge Based System (HKBS) for Software Selection: A Comparative Study. In Proceedings of 2nd International Conference on Emerging Trends in Engineering and Technology, 2009 (pp. 991-997). IEEE.

Jagatic, T. N., Johnson, N. A., Jakobsson, M., & Menczer, F. (2007). Social phishing. Communications of the ACM, 50(10), 94–100. doi.org/10.1145/1290958.1290968

Jensen, M., Dinger, M., Wright, R., & Thatcher, J. (2017). Training to Mitigate Phishing Attacks Using Mindfulness Techniques. Journal of Management Information Systems, 34(2), 597-626.

Johnson, S. D., Aragon, S. R., Shaik, N., & Palma-Rivas, N. (2000). Comparative analysis of learner satisfaction and learning outcomes in online and Face-to-Face learning environments. Journal of Interactive Learning Research, 11(1), 29.

Kalloniatis, A., Ali, I., Neville, T., La, P., Macleod, I., Zuparic, M., & Kohn, E. (2017). The Situation Awareness Weighted Network (SAWN) model and method: Theory and application. Applied Ergonomics, 61, 178–196. doi.org/10.1016/j.apergo.2017.02.002

Kirlappos, I., & Sasse, M. A. (2012). Security education against phishing: A modest proposal for a major rethink. IEEE Security and Privacy Magazine, 10(2), 24–32. doi.org/10.1109/MSP.2011.179

Knapp, T. R. (1978). Canonical correlation analysis: A general parametric significance-testing system. Psychological Bulletin, 85(2), 410–416. doi.org/10.1037/0033-2909.85.2.410

Knofczynski, G. T., & Mundfrom, D. (2008). Sample sizes when using multiple linear regression for prediction. Educational and Psychological Measurement, 68(3), 431–442. doi.org/10.1177/0013164407310131

Kolen, M. J., & Brennan, R. L. (2004). Test equating, scaling, and linking (pp. 201–205). New York: Springer. doi.org/10.1007/978-1-4757-4310-4

Kruger, H. A., Steyn, T., Drevin, L., & Medlin, D. (2008). Password Management: Empirical Results from a RSA and USA Study. In ISSA (pp. 1-11).

Kumaraguru, P. (2009). Phishguru: a system for educating users about semantic attacks. Carnegie Mellon University.

Kumaraguru, P., Sheng, S., Acquisti, A., Cranor, L. F., & Hong, J. (2010). Teaching Johnny not to fall for phish. ACM Transactions on Internet Technology, 10(2), 7. doi.org/10.1145/1754393.1754396

Laribee, L., Barnes, D. S., Rowe, N. C., & Martell, C. H. (2006). Analysis and defensive tools for social-engineering attacks on computer systems. In Information Assurance Workshop, 2006 IEEE (pp. 388-389). doi.org/10.1109/IAW.2006.1652125

Lederman, M. (2012). The 11 laws of likability [electronic resource]: Relationship networking-- because people do business with people they like. New York: American Management Association.

Levene, H. (1960). Robust tests for equality of variances. Contributions to probability and statistics, 1, 278-292.

Lim, C. C., & Jin, J. S. (2006). A study on applying software security to information systems: E-learning portals. IJCSNS, 6(3B), 162.

Luminita, D. C. C. (2011). Security issues in e-learning platforms. World Journal on Educational Technology, 3(3), 153–167.

Luo, X., Brody, R., Seazzu, A., & Burd, S. (2011). Social engineering: The neglected human factor for information security management. Information Resources Management Journal, 24(3), 1–8. doi.org/10.4018/irmj.2011070101

Luo, X. R., Zhang, W., Burd, S., & Seazzu, A. (2013). Investigating phishing victimization with the Heuristic–Systematic Model: A theoretical framework and an exploration. Computers and Security, 38, 28–38. doi.org/10.1016/j.cose.2012.12.003

Magnani, R., Sabin, K., Saidel, T., & Heckathorn, D. (2005). Review of sampling hard-to-reach and hidden populations for HIV surveillance. Aids, 19(Supplement 2). doi:10.1097/01.aids.0000172879.20628.e1

Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. Biometrika, 57(3), 519–530. doi.org/10.1093/biomet/57.3.519

Markowski, C. A., & Markowski, E. P. (1990). Conditions for the effectiveness of a preliminary test of variance. The American Statistician, 44(4), 322–326.

McKay, J., & Marshall, P. (2001). The dual imperatives of action research. Information Technology and People, 14(1), 46–59. doi.org/10.1108/09593840110384771

McKnight, P. E., & Najab, J. (2010). Mann-Whitney U Test. Corsini Encyclopedia of Psychology. doi.org/10.1002/9780470479216.corpsy0524

Melanie Hof .Questionnaire Evaluation with Factor Analysis and Cronbach's Alpha. 2012. Retrieved from http://www.let.rug.nl/nerbonne/teach/rema-stats-meth-seminar/student-papers/MHof-QuestionnaireEvaluation-2012-Cronbach-FactAnalysis.pdf

Merriam, S. (1995). What Can You Tell From An N ofl?: Issues of validity and reliability in qualitative research. PAACE Journal of Lifelong Learning, 4, 50–60.

Mohammad, R. M., Thabtah, F., & McCluskey, L. (2012, December). An assessment of features related to phishing websites using an automated technique. In International Conference for Internet Technology and Secured Transactions (pp. 492-497). IEEE.

Morris, C. N. (1982). On the foundations of test equating. Test equating, 169-191.

Nyíri, K. (2008). The networked mind. Studies in East European Thought, 60(1-2), 149–158. doi.org/10.1007/s11212-008-9044-0

O’brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. Quality and Quantity, 41(5), 673–690. doi.org/10.1007/s11135-006-9018-6

Oh, Y., & Obi, T. (2012). Evaluation of Field Phishing Study Setup Method. International Journal of Information and Network Security, 1(4), 235.

Parrish, J. L., Jr., Bailey, J. L., & Courtney, J. F. (2009). A personality based model for determining susceptibility to phishing attacks. Little Rock: University of Arkansas.

Parrish, J. L., Jr., Bailey, J. L., & Courtney, J. F. (2009). A Personality Based Model for Determining Susceptibility to Phishing Attacks. Little Rock: University of Arkansas.

Pars, C. (2017). PHREE of Phish: The Effect of Anti-Phishing Training on the Ability of Users to Identify Phishing Emails (Master's thesis, University of Twente).

Pedhazur, E. J. (1997). Multiple regression in behavioral research: Explanation and prediction.

Pesonen, M., Kurttila, M., Kangas, J., Kajanus, M., & Heinonen, P. (2001). Assessing the priorities using A’WOT among resource management strategies at the Finnish Forest and Park Service. Forest Science, 47(4), 534–541.

Poepjes, R., & Lane, M. (2012). An information security awareness capability model. ISACM.

Polyu.edu.hk. (2017). Thresholds for interpreting effect sizes. [online] Available at: http://www.polyu.edu.hk/mm/effectsizefaqs/thresholds_for_interpreting_effect_sizes2.html.

Proctor, R. W., Schultz, E. E., & Vu, K. P. L. (2009). Human factors in information security and privacy. Handbook of research on information security and assurance. Hershey, PA: Idea Group Reference.

Ramsey, P. H. (1989). Critical values for Spearman’s rank order correlation. Journal of Educational Statistics, 14(3), 245–253. doi.org/10.2307/1165017

Rice, W. R. (1989). Analyzing tables of statistical tests. Evolution; International Journal of Organic Evolution, 43(1), 223–225. doi.org/10.1111/j.1558-5646.1989.tb04220.x

Rjaibi, N., Rabai, L. B. A., Aissa, A. B., & Louadi, M. (2012). Cyber security measurement in depth for e-learning systems. International Journal of Advanced Research in Computer Science and Software Engineering, 2(11), 107–120.

Rovai, A. P., & Jordan, H. (2004). Blended learning and sense of community: A comparative analysis with traditional and fully online graduate courses. International Review of Research in Open and Distributed Learning, 5(2). doi.org/10.19173/irrodl.v5i2.192

Ruxton, G. D. (2006). The unequal variance t-test is an underused alternative to Student’s t-test and the Mann–Whitney U test. Behavioral Ecology, 17(4), 688–690. doi.org/10.1093/beheco/ark016

Saunders, M. (2012). Choosing research participants. Qualitative organizational research: Core methods and current challenges, 35-52.

Schrammel, J., Köffel, C., & Tscheligi, M. (2009, September). Personality traits, usage patterns and information disclosure in online communities. In Proceedings of the 23rd British HCI group annual conference on people and computers: Celebrating people and technology (pp. 169-174). British Computer Society.

Scott, D. W. (1979). On optimal and data-based histograms. Biometrika, 66(3), 605–610. doi.org/10.1093/biomet/66.3.605

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). Biometrika, 52(3-4), 591–611. doi.org/10.1093/biomet/52.3-4.591

Sheskin, D. J. (2003). Handbook of parametric and nonparametric statistical procedures. CRC Press.

Siegel, S. (1956). Nonparametric statistics for the behavioral sciences.

Singh, A., & Kapoor, B. (2016). Analysis of the Human Factor behind Cyber Attacks.

Sircombe, K. N. (2004). AgeDisplay: An EXCEL workbook to evaluate and display univariate geochronological data using binned frequency histograms and probability density distributions. Computers and Geosciences, 30(1), 21–31.

Sokovic, M., Pavletic, D., & Pipan, K. K. (2010). Quality improvement methodologies–PDCA cycle, RADAR matrix, DMAIC and DFSS. Journal of Achievements in Materials and Manufacturing Engineering, 43(1), 476-483.

Stanton, N. A., Stewart, R., Harris, D., Houghton, R. J., Baber, C., McMaster, R., & Linsell, M. (2006). Distributed situation awareness in dynamic systems: theoretical development and application of an ergonomics methodology. Ergonomics, 49(12-13), 1288-1311.

Steyn, T., Kruger, H. A., & Drevin, L. (2007). Identity theft—Empirical evidence from a Phishing exercise. In New Approaches for Security, Privacy and Trust in Complex Environments (pp. 193-203). Springer.

Two Sample t Test: unequal variances. (n.d.). Retrieved October 29, 2017, from http://www.real-statistics.com/students-t-distribution/two-sample-t-test-uequal-variances/

Trending Sideways. (2017). The Cohen’s d Formula - Trending Sideways. [online] Available at: http://trendingsideways.com/index.php/cohens-d-formula/.

Trochim, W. M., & Donnelly, J. P. (2006). The research methods knowledge base (3rd ed.). Cincinnati, OH: Atomic Dog.

Tukey, J. W. (1977). Exploratory data analysis.

Vasilescu, C., Tatar, E. L., & Codreanu, A. (2011). Integrating Information Security in an E-Learning Environment. In Proceedings of eLearning and Software for Education Conference, 02, 70-75.

Vishwanath, A., Herath, T., Chen, R., Wang, J., & Rao, H. R. (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.

Waclawski, E. (2012). How I use it: Survey monkey. Occupational Medicine, 62(6), 477.

Wallace, D. L. (1959). Simplified beta-approximations to the Kruskal-Wallis H test. Journal of the American Statistical Association, 54(285), 225-230.

Walsh Jr, E., & Brown, L. (2013, March). Connected: Using Skype™ in the College Classroom. In Proceedings of Society for Information Technology and Teacher Education International Conference, 1115-1119.

Wang, M. W., & Stanley, J. C. (1970). Differential Weighting: A Review of Methods and Empirical Studies. Review of Educational Research, 40(5), 663-705. doi:10.3102/00346543040005663

Wang, Q., & Woo, H. L. (2007). Comparing asynchronous online discussions and face-to-face discussions in a classroom setting. British Journal of Educational Technology, 38(2), 272-286.

Watson, G., Mason, Andrew, & Ackroyd, Richard. (2014). Social engineering penetration testing [electronic resource] : Executing social engineering pen tests, assessments and defense. Saint Louis: Elsevier Science.

Williams, E. J., Beardmore, A., & Joinson, A. N. (2017). Individual differences in susceptibility to online influence: A theoretical review. Computers in Human Behavior, 72, 412-421.

Winter, G. (2000). A comparative discussion of the notion of' validity ‘in qualitative and quantitative research. The Qualitative Report, 4(3), 1-14.

Workman, M. (2008). 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.

Wright, M. C., Taekman, J. M., & Endsley, M. R. (2004). Objective measures of situation awareness in a simulated medical environment. Quality and Safety in Health Care, 13(suppl 1), i65-i71.

Wright, R. T., & Marett, K. (2010). The influence of experiential and dispositional factors in phishing: An empirical investigation of the deceived. Journal of Management Information Systems, 27(1), 273-303.

Wright, R. T., Jensen, M. L., Thatcher, J. B., Dinger, M., & Marett, K. (2014). Research Note—Influence Techniques in Phishing Attacks: An Examination of Vulnerability and Resistance. Information Systems Research, 25(2), 385-400.

Woolson, R. F. (2008). Wilcoxon Signed‐Rank Test. Wiley encyclopedia of clinical trials.

Yong, J. (2007). Security modelling for e-Learning. In Proceedings of the 1st International Symposium on Information Technologies and Applications in Education (pp. 1-5).

Zar, J. H. (1972). Significance testing of the Spearman rank correlation coefficient. Journal of the American Statistical Association, 67(339), 578-580.

Zhang, D., Zhou, L., Briggs, R. O., & Nunamaker, J. F. (2006). Instructional video in e-learning: Assessing the impact of interactive video on learning effectiveness. Information and Management, 43(1), 15-27.

Zhang, Y., Hong, J. I., & Cranor, L. F. (2007, May). Cantina: a content-based approach to detecting phishing web sites. In Proceedings of the 16th international conference on World Wide Web (pp. 639-648). ACM.

Zimmerman, D. W. Teacher's Corner: A Note on Interpretation of the Paired-Samples T Test. Journal of Educational and Behavioural Statistics 22.3 (1997): 349-360.

Zimmerman, D. W. (1997). Teacher’s corner: A note on interpretation of the paired-samples t-test. Journal of Educational and Behavioural Statistics, 22(3), 349-360.

Zuev, V. I. (2012). E-learning security models. Journal of Management Information Systems, 7(2), 024-028.