

PHREE of Phish

The Effect of Anti-Phishing Training on the Ability of Users to Identify Phishing Emails

Cas Pars

s1735314

Master Thesis, Business Information Management

University of Twente

Faculty of Behavioral, Management and Social Science

Supervisors:

Professor Dr. M. Junger

Dr. A.A.M. Spil

17 July 2017

ABSTRACT

Phishing attacks evolve and keep on doing harm to victims. Various anti-phishing training techniques have been proposed as a human-oriented solution for phishing. Experimental evaluations show that these training techniques have had mixed success. Therefore, the aim of the present thesis was to develop a new anti-phishing training based on what has been learned from previous research. To achieve this goal this research followed three steps. First, a systematic literature review was conducted: what are the characteristics of anti-phishing training methods that have been published and tested in scientific experiments? Which characteristics of anti-phishing training are central to their effectiveness? Second, the anti-phishing training was developed according to the results of step 1. And third, the anti-phishing training was tested in a randomized controlled trial. The results are summarized as follows.

For the literature review articles were carefully selected according to the Grounded Theory method for rigorously reviewing literature. Articles were only included if they were published in English and address the topic of digital training as a countermeasure for phishing. The review indicated that an effective anti-phishing training has a repetitive, game-based, embedded design in which text is kept simple and short by using a cartoon format. The content of an effective anti-phishing training contains cues to identify phishing emails and phishing URLs as well as a solution for uncertain situations.

Based on these characteristics 'PHREE', a new anti-phishing training, was developed to enhance the ability of users to distinguish phishing emails from legitimate emails. In this game-based training, users play the character of a cartoon called 'Bob Visvanger'. The game contains four levels of anti-phishing training. Each level includes a short and simple instructional video on how to identify phishing emails or phishing URLs and each level ends with four, topic related, practice questions. The training is completed if users pass all four levels of the game.

Subsequently, PHREE was tested in an experiment with 36 participants who were equally and randomly divided over a control group (no training) and an experimental group (PHREE training). Each participant had to identify 10 emails as phishing or legitimate in a pretest, a direct posttest, and a retention test after one week. User' performance was measured by the total amount of correctly identified emails (phishing + legitimate), the amount of correctly identified phishing emails, and the amount of correctly identified legitimate emails. The confidence of users in judging the legitimacy of emails was also measured.

Results indicate that PHREE training improved the ability of users to identify emails (phishing + legitimate) correctly from 68% correct before training to 86% correct after training. PHREE training especially enhanced the ability of users to recognize phishing emails from 52% correct before training to 92% correct after training. User retained this enhanced ability to identify (phishing) emails for at least one week. Trained users performed significantly better than untrained users who identified approximately 72% of all emails (phishing + legitimate) and 59% of the phishing

emails correct at each test moment. PHREE training did not significantly change the confidence of users in their decision-making, nor did it change the ability of users to identify legitimate emails. Finally, results indicated that age and gender had an effect on the amount of correctly identified emails (i.e. older users performed slightly better than younger users and men performed slightly better than women), but education level had no effect.

In conclusion, PHREE strongly enhanced the ability of users to identify (phishing) emails and users retained this ability for at least one week. Overall these pilot test findings strongly support the use of PHREE as a human-oriented solution for phishing. Future research is needed to determine the effect of PHREE in a real-world (corporate) setting.

Kew words: Phishing, Anti-phishing training, Game design, Development and testing

PREFACE

Two years ago, I quitted my fulltime job to start the Master program Business Administration at Twente University. A life decision, that forced me to get out of my safe haven and step into the unknown academic world. This thesis reflects the last phase of my study and concludes my time as a Master student. Doing this research helped me to develop myself academically and personally, as it was the most challenging task I have ever done.

I would like to thank several people who were involved in the process of finalizing this thesis and were of tremendous help. Firstly, I would like to express my great gratitude towards my first supervisor Professor Dr. M. Junger. Thank you for the support, guidance, and fruitful advice, it made me strive for the best. I would also like to thank my second supervisor Dr. A.A.M. Spil for the valuable feedback during the end of the research process.

Secondly, I would like to thank phishing expert and PhD candidate Mr. E. Lastdrager. Your contribution in the formation of the methodology of this research and help with acquiring and formatting phishing emails was indispensable.

Finally, I would like to thank my family, friends, and my girlfriend. Thank you for the endless support, not only during this time of study, but also throughout my entire life.

I hope you enjoy reading this,

Cas Pars

Leiden, 17 July 2017

TABLE OF CONTENT

1. INTRODUCTION.....	7
2. LITERATURE REVIEW OF ANTI-PHISHING TRAINING TECHNIQUES	9
2.1 PHISHING EMAILS	13
2.2 "GOTCHA" EXPERIMENTS	15
2.3 EMBEDDED TRAINING INTERVENTIONS.....	19
2.4 GAME-BASED TRAINING.....	29
2.5 MOST EFFECTIVE TRAINING TECHNIQUES	35
2.6 OTHER PROPOSED ANTI-PHISHING TRAINING METHODS.....	37
2.7 OVERVIEW OF FINDINGS	39
3. DEVELOPMENT AND TEST OF NEW TRAINING MATERIAL.....	40
3.1 DEVELOPMENT OF ANTI-PHISHING TRAINING PHREE	40
3.2 PILOT STUDY: TEST PHREE	45
4. RESULTS	47
4.1 DEMOGRAPHICS	47
4.2 USER PERFORMANCE	47
4.3 USER FEEDBACK	53
5. DISCUSSION AND CONCLUSION	54
5.1 DISCUSSION.....	54
5.2 CONCLUSION.....	57
REFERENCES.....	58
APPENDICES	65
APPENDIX 1: FINAL SELECTION LITERATURE REVIEW IN CHRONOLOGICAL ORDER	65
APPENDIX 2: METHODOLOGICAL DESIGN ANTI-PHISHING TRAINING STUDIES	75
APPENDIX 3: TRAINING CONTENT OF ANTI-PHISHING TRAINING STUDIES	76
APPENDIX 4: STATISTICAL ANALYSIS	77

LIST OF FIGURES

FIGURE 1: TEXT AND GRAPHICS INTERVENTION	20
FIGURE 2: COMIC STRIP INTERVENTION	20
FIGURE 3: PHISHGURU	23
FIGURE 4: TWO-COLUMN TEXT TRAINING	27
FIGURE 5: DESIGN OF PHREE.....	41
FIGURE 6: PROCEDURES OF PHREE	44
FIGURE 7: TOTAL CORRECTLY IDENTIFIED EMAILS	49
FIGURE 8: CORRECTLY IDENTIFIED PHISHING EMAILS.....	50
FIGURE 9: CORRECTLY IDENTIFIED LEGITIMATE EMAILS	52
FIGURE 10: CONFIDENCE OF USERS IN DECISION-MAKING.....	53

LIST OF TABLES

TABLE 1: MOST IMPORTANT DEVELOPMENTS IN ANTI-PHISHING TRAINING RESEARCH.....	10
TABLE 2: CHARACTERISTICS CENTRAL TO THE EFFECTIVENESS OF ANTI-PHISHING TRAINING	40
TABLE 3: METHODOLOGY OF EXPERIMENT	45
TABLE 4: DISTRIBUTION OF EMAILS	46
TABLE 5: GENDER FREQUENCIES.....	77
TABLE 6: AGE FREQUENCIES	77
TABLE 7: EDUCATION LEVEL FREQUENCIES	77
TABLE 8: MEAN DIFFERENCES IN PRETEST SCORES FOR EMAIL SET A, B, AND C.....	77
TABLE 9: CORRELATION BETWEEN GENDER, AGE, EDUCATION AND TCR	78
TABLE 10: TWO-WAY ANOVA INTERACTION EFFECT OF <i>TIME*GROUP</i> ON TCR.....	78
TABLE 11: TWO-WAY ANOVA INTERACTION EFFECT OF <i>TIME*GROUP</i> ON PR.....	78
TABLE 12: TWO-WAY ANOVA INTERACTION EFFECT OF <i>TIME*GROUP</i> ON LR	79
TABLE 13: TWO-WAY ANOVA INTERACTION EFFECT OF <i>TIME*GROUP</i> ON CONFIDENCE	79

1. INTRODUCTION

The number of people using the Internet continues to rise, from 1,03 billion Internet users in 2005 to 3,39 billion in June 2016 ("Internet Users" 2016). On the one hand the Internet brings numerous benefits to its users (Berthon, Pitt, & Watson, 1996; Maignan & Lukas, 1997; Paul, 1996). On the other hand there are downsides of the Internet, as it is a hotspot for hackers, pranksters, and viruses (Paul, 1996). From a financial perspective, recent and increasing problems emerge as a result of phishing (Kumaraguru, Rhee, Acquisti, et al., 2007). Phishing is defined as: "*a scalable act of deception whereby impersonation is used to obtain information from a target.*" (Lastdrager, 2014, p. 8).

Phishing has devastating consequences for firms and individuals (Emigh, 2005; Hong, 2012). The total estimated damage in direct losses for individuals ranges between 61 million and 3 billion a year in the USA (Hong, 2012). The average direct costs for companies are estimated at \$320 million per year (Anderson et al., 2012). Besides, a recent report about phishing showed a tremendous increase in the number of phishing attacks, from 48 thousand in October 2015 to 123 thousand in March 2016. Although, a seasonal increase of phishing is standard, a 250 percent increase that continues until March 2016 is reason for concern (Aaron & Manning, 2016).

Three main reasons for the continuous problem of phishing exist. Firstly, technical solutions for phishing have been developed but cannot prevent that attacks reach users (Forte, 2009; Hong, 2012; Zhang, Egelman, Cranor, & Hong, 2007). Secondly, despite concerns about online privacy and security (Sheehan & Hoy, 2000) users trust websites and are willing to give personal information (Downs, Holbrook, & Cranor, 2006; Milne & Gordon, 1993; Sheehan & Hoy, 2000). Consequently, users are vulnerable for phishing (Aloul, 2010; Jagatic, Johnson, Jakobsson, & Menczer, 2007). Thirdly, research shows that users fall for phishing because they lack the knowledge to protect themselves (Aburrous, Hossain, Dahal, & Thabtah, 2010; Mohebzada, Zarka, Bhojani, & Darwish, 2012). It is argued that training is necessary to increase users' knowledge and so their ability to identify and avoid phishing emails (Mohebzada et al., 2012; Steyn, Kruger, & Drevin, 2007; Tyler, 2016).

The effect of anti-phishing training has been tested in many studies while using a variety of content and design features. While some studies tested the effect of a simple warning message (e.g. Bowen, Devarajan, & Stolfo, 2011), others investigated more sophisticated training designs as cartoons (e.g. Gupta & Kumaraguru, 2014) or games (e.g. Sheng et al., 2007). Besides, where some studies tested training techniques that provided one cue to identify and avoid phishing attacks (e.g. Alnajim & Munro, 2009a), others tested training techniques that included many cues (e.g. Kumaraguru, Sheng, Acquisti, Cranor, & Hong, 2010). The two most tested training techniques are PhishGuru, a program that educates users about phishing during their regular use of email (Kumaraguru, Rhee, Sheng, et al., 2007), and Anti-Phishing Phil, a game that teaches users how to identify phishing URLs (Sheng et al., 2007). However, current literature show that anti-phishing

training has had mixed success (e.g. Caputo, Pfleeger, Freeman, & Johnson, 2014). Therefore, the goal of the present thesis is to develop an anti-phishing training based on a systematic literature review and test its effectiveness in enhancing the ability of users to identify phishing emails. To achieve the purpose, this thesis follows three steps:

1. The relevant literature is reviewed: what are the characteristics of anti-phishing training methods that have been published and tested in scientific experiments? Which characteristics of anti-phishing training are central to their effectiveness?
2. According to the findings of step 1, a new anti-phishing training is developed.
3. The anti-phishing training is tested in a randomized controlled pilot experiment.

This thesis makes three important contributions. First, this study is the first to present an overview of characteristics that are central to the effectiveness of anti-phishing training. This scientific contribution is useful for researchers, but definitely also for managers that want to train their employees, but do not know how, when, or what to teach. Second, an overview of current anti-phishing training literature provides researchers and companies with a clear indication of what benefit anti-phishing training may bring to its users. And third, this thesis presents the results of a complete new anti-phishing training, developed and tested in a scientific pilot experiment.

The thesis outline is as follows. Chapter 2 provides the literature review. Chapter 3 presents the development and test of the new anti-phishing training. Then, in chapter 4 the results of the experiment are displayed. Finally, chapter 5 presents the discussion and conclusion.

2. LITERATURE REVIEW OF ANTI-PHISHING TRAINING TECHNIQUES

This chapter describes the findings of previous studies regarding the effect of anti-phishing training according to the Grounded Theory method for rigorously reviewing literature (Wolfswinkel, Furtmueller, & Wilderom, 2013).

First the scope of the review was determined as well as the inclusion and exclusion of criteria. This literature review only includes articles that were published in English and that address the topic of digital training as a countermeasure for phishing. Conversely, studies were excluded if they were not published in English, not focused on phishing or training, or if they proposed technical countermeasures. Subsequently, computer science was considered as the main field for this review since anti-phishing training experiments are published in computer science journals. The learning science field was excluded in the literature review, as the goal of this study was to learn from previous anti-phishing experiments.

Due to its accessibility, the literature review started with a search through the Scopus database, as this index ensures most articles on phishing as they include journals as Computer and Security, Acm Transaction on Internet Technology, and IEEE Security and Privacy. Web of Science, the other database supported by the University of Twente, was not considered, as it did not bring up any additional useful articles. Literature was found by using the search words "Phishing" and "Training" in both the title and the abstract. In total 116 articles (journal articles or conference papers) were found. However, this sample lacked information on what content anti-phishing training should have. To address this, a second search in Scopus was performed to find papers with the words "Sensitive information" in the abstract and "Phishing" in the title. The second search resulted in 62 articles. To make sure no information was missed, synonyms for training were also used to find relevant articles. A search with the words "Phishing" and "Educating" in the title or abstract, while not including "Training" or "Sensitive information", resulted in 37 articles. Finally, a search on "Phishing" and "Learning" in the title or abstract, while not including "Training", "Sensitive information", or "Educating" resulted in 71 articles. Therefore, in total 286 papers were found.

From this pool of articles, only the relevant papers were included. Therefore, first 23 doubles were filtered out. Then, the abstracts of the papers were scanned. Most articles proposed technological countermeasures (e.g. Bergholz et al., 2010; Falk & Kucherawy, 2010; He et al., 2011; Smadi, Aslam, Zhang, Alasem, & Hossain, 2015; Xiang, Hong, Rose, & Cranor, 2011), other studies did not explicitly focus on phishing (e.g. Claffey Jr & Regan, 2011; Song, Yang, & Gu, 2010; Stikic, Berka, & Korszen, 2015), or could not be related to training (e.g. Albladi & Weir, 2016; Norris, Joshi, & Finin, 2015; Welk et al., 2015). In line with the scope of this thesis, these articles were excluded. Papers were also excluded if they aimed to show the need for training but did not perform any tests (e.g. Aloul, 2010; Tyler, 2016), presented training as part of a larger anti-phishing model (e.g. Besimi, Shehu, Abazi-Bexheti, & Dika, 2009; Frauenstein & Von Solms, 2014), or merely sketched the

profile of phishers (e.g. Aston, McCombie, Reardon, & Watters, 2009; Halaseh & Alqatawna, 2016) or its victims (e.g. Flores, Holm, Nohlberg, & Ekstedt, 2015; Frauenstein & Von Solms, 2014).

From the 286 papers, 246 articles dropped out, as their content was not relevant. Seven articles did not show up via the various search terms but were added as a result of a citation search using Google Scholar (Alnajim & Munro, 2009b; Clark & Mayer, 2016; Dhamija, Tygar, & Hearst, 2006; Jansson & Von Solms, 2011; Kearney & Kruger, 2014; Smith, Papadaki, & Furnell, 2009; Yang, Tseng, Lee, Weng, & Chen, 2012). So the final sample for the literature review contained 47 articles. These 47 papers were studied thoroughly to understand the findings within each article fully. Analytical tables were built up to compare the outcomes between papers (appendix 1 and appendix 2). These tables contain essential information on year of publication, authors, methodology, and main results. The articles were put in chronological order to follow the development of anti-phishing training materials. By using the knowledge gained from analyzing and comparing these articles, it was possible to define characteristics central to the effectiveness of anti-phishing training. The results of the most important anti-phishing studies are summarized in table 1. Table 1 also serves as a guideline throughout the rest of this chapter in which each study (design, results, limitations) and its specific terminologies are described in detail.

Table 1: Most Important Developments in Anti-Phishing Training Research

Author	Development	Results	Chapter
(Dodge Jr, Carver, & Ferguson, 2007)	First experiment in which unknowing users were sent simulated phishing emails.	<ul style="list-style-type: none"> • Sending simulated phishing emails to unknowing users enhanced the ability of users to identify phishing emails. 	2.2
(Alnajim & Munro, 2009a)	First experiment in which unknowing users were shown a warning message after they tried to fill out information on a simulated phishing websites.	<ul style="list-style-type: none"> • Presenting a warning message after users tried to submit information on a phishing website enhanced the ability of users to identify phishing emails. • The website warning messages had a greater effect on the ability of users to identify phishing emails than sending anti-phishing tips via email. 	2.2
(Aburrous et al., 2010)	Tested the effect of experience with phishing on the ability of users to identify websites as legitimate or phishing.	<ul style="list-style-type: none"> • Experience with phishing enhanced the ability of users to identify phishing websites from legitimate websites. 	2.2

Author	Development	Results	Chapter
(Kumaraguru, Rhee, Acquisti, et al., 2007)	First embedded training interventions in which users were shown a training message after they clicked on a simulated phishing email.	<ul style="list-style-type: none"> • Falling for simulated phishing emails enhanced motivation to learn. • Embedded training interventions enhanced the ability of users to identify phishing emails. • Embedded training messages had a greater effect on the ability of users to identify phishing emails than security notices. • A comic strip intervention had a greater effect on the ability of users to identify phishing emails than a text (and image) intervention. 	2.3
(Kumaraguru, Rhee, Sheng, et al., 2007)	Developed PhishGuru, an embedded training intervention.	<ul style="list-style-type: none"> • Embedded training interventions had a greater effect on the ability of users to identify phishing emails than non-embedded training emails. • Users retained knowledge gained by PhishGuru up to seven days. 	2.3
(Kumaraguru, Sheng, Acquisti, Cranor, & Hong, 2008)	First real-world corporate test with PhishGuru and with spear training.	<ul style="list-style-type: none"> • Embedded training interventions enhanced the ability of users to identify phishing emails in a real-world corporate setting. • Trained users could retain their knowledge up to seven days in a real-world corporate setting. • Keeping text in anti-phishing training simple and short seemed an effective way to enhance the ability of users to identify phishing emails. 	2.3
(Kumaraguru, Cranshaw, et al., 2009)	First multiple-training experiment with PhishGuru, and to test knowledge retention after 28 days.	<ul style="list-style-type: none"> • Users in a single-training (trained on day 0) condition could retain the ability to identify phishing emails up to 28 days. • Users in a multiple-training (day 0 and day 14) condition were better able to identify phishing emails at day 16 and day 21, but there was no significant difference at day 28. 	2.3

Author	Development	Results	Chapter
(Caputo et al., 2014)	Developed a two-column text training, an embedded training intervention.	<ul style="list-style-type: none"> • Employees that received the two-column text training did not perform significantly better than employees in the control condition in identifying phishing emails. • Multiple reasons for this outcome are: (1) the training was not good. (2) Participants did not read the training. (3) The control group also received an embedded warning message. (4) There was no direct posttest, only a second test that was performed months after the first test. 	2.3
(Gupta & Kumaraguru, 2014)	Tested an Anti-Phishing Landing Page, an embedded website intervention.	<ul style="list-style-type: none"> • Users clicked less often on blacklisted websites after they saw the Anti-Phishing Landing Page. 	2.3
(Sheng et al., 2007)	Developed Anti-Phishing Phil, the first published game-based anti-phishing training.	<ul style="list-style-type: none"> • Playing Anti-Phishing Phil enhanced the ability of users to identify phishing URLs. • Playing Anti-Phishing Phil had a greater effect on the ability of users to identify phishing URLs than existing training materials (eBay and Microsoft tutorials). • Seeing the lessons provided in Anti-Phishing Phil printed out on paper did not have a greater effect on the ability of users to identify phishing URLs than existing training materials (eBay and Microsoft tutorials). 	2.4
(Kumaraguru et al., 2010)	Knowledge retention test with Anti-Phishing Phil.	<ul style="list-style-type: none"> • Users who played Anti-Phishing Phil and scored poorly in the pretest improved their ability to identify phishing websites significantly after training and retained this knowledge for one week. 	2.4
(Sercombe & Papadaki, 2012)	Developed the Malware Man game, a game-based training.	<ul style="list-style-type: none"> • Trained users were better in answering survey questions about phishing than untrained users. 	2.4

Author	Development	Results	Chapter
(Yang et al., 2012)	Developed the Anti-Phishing Education Game, a game-based training.	<ul style="list-style-type: none"> • The Anti-Phishing Education game significantly enhanced the ability of users to identify phishing websites. • Users in the control group (no training) also significantly enhanced their ability to identify phishing websites. 	2.4
(Canova, Volkamer, Bergmann, & Reinheimer, 2015)	Developed NoPhish, a game-based training. The experiment included a retention test after five months.	<ul style="list-style-type: none"> • NoPhish statically significant enhanced the ability of users to identify phishing URLs directly after training. • After five months, users still performed significantly better than before training but significantly worse than directly after training. 	2.4
(Dodge, Coronges, & Rovira, 2012)	Tested the difference between presenting an error message, feedback, or training after users fall for phishing.	<ul style="list-style-type: none"> • After 10 days there was no significant difference between the three treatment groups in their ability to identify phishing emails. • After 63 days the ability to identify phishing emails was the highest for trained users, than for users that received feedback, and the lowest for users that received an error message. 	2.5
(Mayhorn & Nyeste, 2012)	Combined game-based training (Anti-Phishing Phil) with embedded training interventions (cartoons).	<ul style="list-style-type: none"> • Directly after training, trained users performed significantly better than a control group (no training) in identifying phishing emails. • The positive effect of the training remained in the second test, however, this time not statistically significant different from the control group. 	2.5

The chapter outline is as follows. First, the characteristics of phishing emails are described in chapter 2.1. Second, in chapter 2.2 until 2.6 gotcha experiments, embedded training interventions, game-based training, and other anti-phishing training techniques are discussed. Finally, in chapter 2.7 the main findings of the literature review are summarized.

2.1 Phishing Emails

Phishing is initiated via several instruments; a very popular method is phishing via email (Kumaraguru, Rhee, Acquisti, et al., 2007). These phishing emails try to trick users into giving personal information or to click on links to phishing websites. A wide range of tactics to trick users into giving personal information is used (Kumaraguru, Rhee, Acquisti, et al., 2007). The phishing email, for instance, may ask users to verify their bank account, update their password or to send a

small amount off money to a charity foundation in Africa. Nevertheless, phishing emails have certain characteristics.

Characteristics of phishing emails without links. Aggarwal, Kumar, and Sudarsan (2014) examined features of phishing emails that aim to get potential victims' information by luring them into replying to phishing emails. Aggarwal et al. (2014) exploited the common features within such emails by analyzing 600 phishing emails without links over a period of six months. They found six characteristics of phishing emails without links. Firstly, people who sent phishing emails began this process by finding email lists on websites. Subsequently, they sent phishing emails to the entire list. One indicator of these emails was that they come without the name of the recipient (Aggarwal et al., 2014). Secondly, most of the examined phishing emails promised an amount of money in some way or the other so that the potential victim was tempted to respond to the email (Aggarwal et al., 2014). Thirdly, the phishers used some sort of reasoning (story line) for the victim to believe that the intention of the email was legitimate. Fourthly, the emails often asked for personal sensitive information (Aggarwal et al., 2014). Fifth, the phishing emails without links often ended with a sentence that requested the victim to reply to a particular email address (Aggarwal et al., 2014). Most of the time the sender's email address was different form the reply-to email address (Aggarwal et al., 2014). Sixth, this reply-luring request often contained a sense of urgency meant to let the victim reply as soon as possible. According to Aggarwal et al. (2014) reasons for the urgency request were: (1) it gives victims less time to think logically, and (2) when victims reply the email it is considered as non-spam and therefore, the chance that the email will be blocked or blacklisted is reduced (Aggarwal et al., 2014). A blacklisted email is an email that has officially been classified as phishing (Alnajim & Munro, 2009a). Finally, in the 600 analyzed emails Aggarwal et al. (2014) found no pattern in the way attackers made victims believe that the email was legitimate.

Characteristics of phishing emails (with links). According to Downs et al. (2006) users should treat emails with suspicion for phishing when an email asks to follow a link to update account information, or when an email threatens with consequences for not immediately providing personal information. Emails that come from organizations with which the user does not have an account should also be treated with suspicion. Another reason for skepticism is when the email claims to be from an organization but it contains misspelled words, odd spacing, or sloppy grammar (Downs et al., 2006). A final reason for suspicion is when the senders' address in the "From" field is different from than the name usually used by the company (Caputo et al., 2014).

Most phishing emails contain a request for personal information, either directly or via a link to phishing websites in the email (Downs et al., 2006). A characteristic of a phishing email is that it often contains a link to a phishing URL (Kumaraguru, Rhee, Acquisti, et al., 2007). Users can examine the URLs behind these links, without clicking on them, by hovering over the link with the mouse (Downs et al., 2006). Examining these links will then show the attached URL.

-
1. *Phishing emails often request for personal information.*
 2. *Phishing emails often contain a sense of urgency.*
 3. *Phishing emails often have a mismatch between the senders' email address in the "From" field and the company name or reply-to email mentioned in the body of the email.*
 4. *Phishing emails often contain a threat to stimulate a response.*
 5. *Phishing emails often contain misspelled words, odd spacing, or sloppy grammar.*
 6. *Phishing emails often contain links to phishing websites.*
 7. *Hovering the mouse over a link in an email will reveal the linked URL.*
-

Sophisticated phishing emails. While many phishing emails are plagued with poor grammar, it is expected that phishers start using proper grammar in the future (Marett & Wright, 2009). So what are tactics of deception detection for more sophisticated phishing emails? First, phishing emails may use a name that is known to the receiver in the body of the email, for example by including the name of a colleague (Marett & Wright, 2009). Second, phishers distract people from what is really going on by personalizing the email. One way to do this is by spear phishing (Marett & Wright, 2009). The difference between spear phishing and general phishing is that spear phishing is addressed directly to the victim and uses inside information. A general phishing attack is often less focused on one victim and not addressed to the victim personally, but rather aims at a broad public. As a result spear phishing is more effective and needs far fewer attacks to achieve the same financial benefits as general phishing attacks (Caputo et al., 2014). Finally, phishers mimic official emails so it appears to be legitimate. Phishers, for example, create email accounts (visible in the "From" field) that look like the email accounts from official organizations (Marett & Wright, 2009).

(Spear) phishing emails can look very similar to legitimate emails.

2.2 "Gotcha" Experiments

The first human-oriented anti-phishing experiments did not include training (Dodge Jr et al., 2007). Rather, these studies tested the effect of a "gotcha" moment. A "gotcha" moment emerges when users are sent simulated phishing messages in the context where they would normally be attacked as part of a test. For example when employees receive simulated phishing emails in their corporate email inbox. The idea is that when users fall for these phishing attacks they realize how vulnerable they are and, therefore, act more careful in the future.

Error message. The first "gotcha" experiments were performed with students from the United States Military Academy (Dodge Jr et al., 2007). The unknowing students were sent simulated phishing emails to their regular school email to determine the efficacy of the academy's user security training. Four types of phishing emails were used. The first type asked users to click on a link to see their grade report. The second type was identical to the first type except that it asked students to open

an attachment. In the third email type students were asked to click on a link that forwarded them to a website that requested for their social security number. Finally, the fourth type asked students to click on a link to download and run an application (Dodge Jr et al., 2007). The emails were presented in a way that they were questionable enough to raise suspicion. Dodge Jr et al. (2007) performed three tests, a pilot test included 515 students, the second test 4,118 students, and the third test was performed with 4,136 students.

If students fell for the trap by clicking on a link or attachment in one of the simulated phishing emails they saw an error message (Dodge Jr et al., 2007). So students were not trained nor informed about why they received phishing email, or how they could have identified the email as phishing.

The failure rate for the pilot test was 80%, and approximately 40% for the two subsequent experiments. The average failure rate per email type (over the three experiments) was 38% for students that received the link to grade report email, 50% for students that received the grade report attachment email, and 46% for the social security number email (Dodge Jr et al., 2007). The fourth email type was excluded in the analyses due to technical difficulties.

When analyzing the failure rate per class it was found that freshman students (more than 50%) fell for phishing more often than seniors (less than 20%), indicating that the longer a student was at the Military Academy, receiving annual cyber security training, the lower the chance they fell for phishing (Dodge Jr et al., 2007). Two classes participated in three phishing experiments within the same year. For one class, the failure rate dropped from 84% during the first test, to 44% at the second test, and to 24% at the last test. For the other class the failure rate dropped from 91% at test one, to 39% at test two, and to 30% at test three (Dodge Jr et al., 2007).

The study concluded that students kept on disclosing personal information that should not have been disclosed (Dodge Jr et al., 2007). On the bright side, the study showed that with the iteration of the exercise of sending simulated phishing emails, the amount of victims reduced (Dodge Jr et al., 2007).

A conclusion that was later confirmed by Aburrous et al. (2010) who found that experience with phishing enhanced the ability of users to recognize phishing websites. They compared employees that were confronted with phishing before ($n = 50$) to employees that had no experience with phishing ($n = 50$) in identifying phishing websites. It was found that the employees with experience identified 72% of the 50 presented websites correctly, while users without experience identified 28% of the websites correctly (Aburrous et al., 2010).

Tricking users with simulated phishing emails seems to be an effective way to enhance the ability of users to identify phishing emails.

Warning message. The error message that Dodge Jr et al. (2007) presented after users fell for phishing was replaced by a warning message in later research (e.g. Bowen et al., 2011). For the

purpose of this study a warning message is defined as a one-time text training that warns users for phishing and provides a maximum of three tips or tricks to identify and avoid phishing attacks, but does not include graphics, oral explanations, sounds, examples, or test questions. Three studies examined the effect of sending simulated phishing attacks in combination with a warning message (Alnajim & Munro, 2009a; Bowen et al., 2011; Jansson & Von Solms, 2013).

Falling for phishing emails. (1) In a study performed by Jansson and Von Solms (2013) 25,579 unknowing students from the Nelson Mandela Metropolitan University in South Africa were sent two simulated phishing emails over a period of two weeks. Both emails invited students to react in an insecure way. Insecure meant, in this study, that users responded with filling out private information or when they downloaded an exe file. Users that reacted insecurely in the first cycle received a red-screen warning informing them of their insecure behavior (Jansson & Von Solms, 2013). Besides, users received an email message with attachment. The email made students aware of their insecure behavior in more detail and the attachment provided the tip: "*do not open files in unexpected emails*" (Jansson & Von Solms, 2011, p. 77).

Comparing individual results between the two cycles made it possible to measure improvement. During the first cycle, 14% of the active email users (1,304 people out of 9,273) reacted insecure, by the second cycle, this percentage dropped to 8% (664 people out of 8,231) (Jansson & Von Solms, 2013). Based on the difference in active users during the two cycles, there were 42.63% less reactions in the second cycle than there were in the first cycle (Jansson & Von Solms, 2013). In total 976 users fell for phishing in week one, but not in week two, while being active email users in both weeks. So 11.85% of the total population learned from the first attack (Jansson & Von Solms, 2013). For this reason the study concluded that sending simulated phishing emails in combination with warning messages can positively influence secure email behavior (Jansson & Von Solms, 2013).

(2) A study performed by Bowen et al. (2011) included multiple rounds of simulated phishing emails. During the first round 500 students and staff members from the Columbia University were sent simulated phishing emails. Only users that fell victim in round one were selected for the next round a few weeks later, in which they received a variation of the first phishing email. This process continued until all students identified and avoided the phishing attacks. Afterwards, the experiment was repeated with a population of 2,000 students (Bowen et al., 2011).

Every time users fell for a simulated attack, regardless of the round they were in, they were presented the following warning message:

The Columbia University IDS Lab is conducting experiments designed to measure the security posture of large organizations and to educate users about safe practices so that they avoid falling prey to malicious emails. The emails automatically generated and sent to users of Columbia's network and email system are designed to test whether users violate basic security policies. Although our emails are completely benign, please be aware that many email are sent that are

designed to trick unsuspecting users into giving up identity information (Bowen et al., 2011, p. 232).

The results of the experiment showed that both the first ($N = 500$) and the second ($N = 2,000$) experiment were repeated until the fourth phishing email. At the first experiment 313 users fell for the first phishing email, from these 313 users 21 users fell for the second phishing email, from these 21 users only one user fell for the third phishing email, and no one fell for the fourth phishing email (Bowen et al., 2011). At the second experiment there were 384 victims in round one, 29 victims in round two, four victims in round three, and no victims in round four (Bowen et al., 2011). This showed again that sending simulated phishing emails in combination with a warning message enhanced secure behavior.

Falling for phishing websites. One study made use of simulated phishing websites in combination with a warning message (Alnajim & Munro, 2009a) Phishing is not restricted to email. Phishing messages are, for example, also sent via social media, and in many cases phishing messages contain links to phishing websites (Kumaraguru, Rhee, Acquisti, et al., 2007). The "gotcha" technique to makes users feel vulnerable for phishing can also be applied to those websites.

(3) Alnajim and Munro (2009a) were the first to develop such a program (APTIPWD). In APTIPWD a warning was presented after users tried to submit information on a blacklisted website. A blacklisted website is a website that has officially been classified as phishing (Alnajim & Munro, 2009a). The APTIPWD program would present users the following message:

A fake website's address is different from what you are used to, perhaps there are extra characters or words in it or it uses a completely different name or no name at all, just numbers. Check the True URL (Web Address). The true URL of the site can be seen in the page 'Properties' or 'Page Info': While you are on the website and using the mouse Go Right Click then Go 'Properties' or 'Page Info'. If you don't know the real web address for the legitimate organization, you can find it by using a search engine such as Google (Alnajim & Munro, 2009a, p. 406).

The program was tested in a laboratory setting with 36 participants that had no technical knowledge (Alnajim & Munro, 2009a). The participants were asked to interact with an email inbox that belonged to an imaginary "Dave Smith", an employee. In total the email inbox contained 14 emails (phishing or legitimate) from which the eighth was a training email (Alnajim & Munro, 2009a). If users clicked on the link in this training email, they proceeded to the linked phishing website. Only if users tried to submit personal information on this blacklisted website (by clicking on the submit button) they saw the warning message (Alnajim & Munro, 2009a). To test the effect of training on the ability to identify emails, the results of a control group (saw a regular email) and two

experimental conditions were compared. The experimental conditions consisted out of a new approach condition (APTIPWD) and an old approach condition (anti-phishing tips via email).

The study showed that untrained users identified 52% of the emails correctly (as phishing or legitimate) in both parts of the experiment. On the one hand, users in the old approach identified 50% correctly before they received the email with anti-phishing tips and 52% correctly afterwards. Therefore, users in the control group and users in the old approach condition did not significantly improve in the second part of the experiment as compared to the first part. On the other hand, users in the new approach condition estimated 52% of the websites correctly before the APTIPWD warning message (similar to the other treatment groups), while after the warning message 77% of the websites (significantly better than the other treatment groups) were correctly identified (Alnajim & Munro, 2009a).

Tricking users with simulated phishing attacks followed by a warning message seems to be an effective way to enhance the ability of users to identify phishing emails.

2.3 Embedded Training Interventions

Just like the "gotcha" experiments, embedded training uses the design of sending simulated phishing attacks to unknowing users in the context where they would normally be attacked (Kumaraguru, Rhee, Acquisti, et al., 2007). Additionally, if users fall for the attack (for example by clicking on a link) they are presented training interventions (Kumaraguru, Rhee, Acquisti, et al., 2007). The idea of embedded training intervention is to motivate users for anti-phishing training by showing how vulnerable they are (Kumaraguru, Rhee, Sheng, et al., 2007). In this study embedded anti-phishing training is defined as anti-phishing training that is initiated immediately after users fall for simulated phishing attacks.

A training intervention is defined as a one-page training that warns users for phishing and provides a minimum of four tips or tricks to identify and avoid phishing attacks, and can include graphics, oral explanations, sounds, examples, or practice questions (figure 1). These extra tips, as compared to the earlier discussed warning messages (maximum three tips to avoid phishing), may enhance phish avoidance behavior. Because even if users are aware of phishing, they do not link this awareness to useful strategies to avoid phishing attacks (Downs et al., 2006). Six studies examined the impact of embedded training interventions (Caputo et al., 2014; Gupta & Kumaraguru, 2014; Kumaraguru, Cranshaw, et al., 2009; Kumaraguru, Rhee, Acquisti, et al., 2007; Kumaraguru, Rhee, Sheng, et al., 2007; Kumaraguru et al., 2008).

Comic strip intervention. Kumaraguru, Rhee, Acquisti, et al. (2007) were the first to test the effect of an embedded training intervention. To do this they designed a text and graphics intervention (figure 1) and comic strip intervention (figure 2).

Figure 1: Text and Graphics Intervention (Kumaraguru, Rhee, Sheng, et al., 2007, p. 5)

Protect yourself from Phishing Scams

Clicking on links within emails like the one in the "amazon.com" email you've just read puts you at risk for identity theft and financial loss. This email and tutorial were developed by Carnegie Mellon University to teach you how to protect yourself from these kind of phishing scams.

1. What's a phishing scam?

- Scammers send fake emails impersonating well-known companies to trick you into giving them your personal information.
- Giving up your personal information such as Social Security Number, credit card number, or account password will lead to identity theft and financial loss.

2. What does a phishing scam look like?

3. What are simple ways to protect yourself from phishing scams?

- Never click on links within emails:** Never click on links within emails or reply to emails asking for your personal information.
- Initiate contact:** Always access a website by typing in the real website address into the web browser.
- Call customer service:** Never trust phone numbers within emails. Look it up yourself and call the customer service when email seems suspicious.
- Never give out personal information:** Never give out personal information upon email request. Companies will rarely ask for your personal information via emails.

PHISHING SCAM EXAMPLE

Subject: Revision to Your Amazon.com Information
 From: "Amazon" <service@amazon.com>
 Date: Tue, April 11, 2006 4:04 pm
 To: bsmith@cognix.com
 Priority: Normal
 Options: View Full Header | View Printable Version

amazon.com

At the last reviewing at your amazon account we discovered that your information is inaccurate. We apologize for this but because most frauds are possible because we dont have enough information about our clients, we require this verification. Please login and reenter you're personal information.

Please follow this link to update your personal information:
<http://www.amazon.com/exec/obidos/sign-in.html>

(To complete the verification process you must fill in all the required fields)
<http://www.amazonaccount.net/exec/obidos/flex-sign-in.htm?104-2497720-5229511>

Professional & legitimate looking design

Urgent messages

Account status threat

Links don't match with status bar when mouse is moved over

Figure 2: Comic Strip Intervention (Kumaraguru, Rhee, Sheng, et al., 2007, p. 5)

Protect Yourself from Phishing Scams

Clicking on links within emails like the one in the "amazon.com" email you've just read puts you at risk for identity theft and financial loss. This email and tutorial were developed by Carnegie Mellon University to teach you how to protect yourself from these kind of phishing scams.

SCAMMER PLANS ATTACK...
 I CAN MAKE A PROFESSIONAL & LEGITIMATE LOOKING EMAIL IMPERSONATING A WELL-KNOWN COMPANY.

I'LL FORGE THE SENDER'S ADDRESS TO LOOK GENUINE
 From: service@amazon.com

I'LL THREATEN USER'S ACCOUNT STATUS WITH URGENT MESSAGE
 Your account will be suspended if you don't update your info.

I'LL INCLUDE A DISGUISED LINK WITHIN THE EMAIL
<http://www.amazon.com/update>

NOW I'LL SEND THIS EMAIL TO MANY USERS
 To: Amazon Member
SEND (click!)

USER RECEIVES SCAM...
 LET'S CHECK WHAT THE NEW EMAIL IS ABOUT
 YOU'VE GOT NEW MAIL!

IT'S ASKING FOR MY ID & PASSWORD, AND LINK LOOKS SUSPICIOUS! I NEVER CLICK ON LINK WITHIN EMAILS
 From: service@amazon.com
 Subject: Revision to Your Account
<http://www.amazon.com/updates> (NOT SAME)
<http://amazon-link.net/account>

I'LL TYPE IN AMAZON.COM IN A NEW BROWSER
 < > G X P
<http://www.amazon.com>

I'LL FIND & CALL REAL CUSTOMER SERVICE CENTER
 1-800-XXX-XXXX

I'LL NEVER GIVE UP MY PERSONAL INFORMATION UPON EMAIL REQUEST
 Username: [redacted]
 Password: [redacted]
 SSN: [redacted]
 Credit Card Number: [redacted]

I WILL NEVER ALLOW SCAMMERS TO STEAL MY PRECIOUS IDENTITY!

The two training interventions had a similar content. Users were taught that criminals could make emails that look like legitimate emails from organizations. Phishers would do this by forging the sender and the link in the email to look genuine. Users were also taught that phishing emails often include a threat to reply on the message urgently (Kumaraguru, Rhee, Acquisti, et al., 2007). Then, based on an analyses of 25 online anti-phishing tutorials, users were instructed to: "(1) never click on links within emails, (2) type in the website address into the web browser (3) find and call a real

customer service, (4) never give out personal information" (Kumaraguru, Rhee, Acquisti, et al., 2007, p. 5). The rationale for never click on links in emails was that it is difficult for non-experts to distinguish between a phishing link and a legitimate link. The rationale for manually typing the URL was that phishing URLs appear to be genuine URLs but are not identical. The rationale for calling customer service (look up the number via a trusted source like the Yellow Pages) was that companies could tell the user if they sent email. Finally, the rationale for never give personal information was that companies rarely ask for such information (Kumaraguru, Rhee, Acquisti, et al., 2007).

The designs of the two embedded training interventions deferred slightly. The text and graphics intervention showed a screenshot of a phishing email and explained in text how users could identify and avoid phishing attacks (figure 1). The comic strip intervention was presented in a comic strip format and, therefore, contained less textual information (figure 2).

To test their interventions Kumaraguru, Rhee, Acquisti, et al. (2007) recruited 30 participants with little technical knowledge by handing out flyers around the Carnegie Mellon University and local neighborhoods in the USA. The 30 participants were divided into equal groups that represented a text and graphics intervention condition, a comic strip intervention condition, and a security notice condition. Kumaraguru, Rhee, Acquisti, et al. (2007) described the security notices as typical security emails sent out by companies to warn users about phishing.

For the experiment Kumaraguru, Rhee, Acquisti, et al. (2007) simulated a working environment by giving participants the role of "Bobby Smith" a business administrator for Cognix Inc. Participants would sit at a desk in a laboratory and had to imagine that the desk they were sitting at was Bobby's office desk (Kumaraguru, Rhee, Acquisti, et al., 2007). Subsequently, they showed each participant Bobby's email inbox and asked them to process and react to the emails as they would normally do at their job (Kumaraguru, Rhee, Acquisti, et al., 2007). The inbox contained 19 emails from which the third, fourteenth, sixteenth, and seventeenth were phishing emails and the fifth and eleventh were training emails. Users did not know they were participating in a study about phishing, and the anti-phishing training interventions were unannounced. Hence, the experimental setup made it possible to test embedded training in a laboratory setting.

Security notices. On the one hand results showed that sending security notices was not an effective way to teach users about phishing attacks. Only five users (50%) clicked on the link in the first security notice training email to learn about phishing. Among these five users two users actually read the training materials, whereas the other three quickly skimmed the training materials and closed the training window (Kumaraguru, Rhee, Acquisti, et al., 2007). Besides, 90% of the users in the security notice group fell for the first phishing email and 90% of the users fell for the final phishing email (Kumaraguru, Rhee, Acquisti, et al., 2007). Moreover, the mean percentage of users that fell for phishing over the last three attacks was 63%.

Sending out security emails seems not an effective way to enhance the ability of users to identify phishing emails.

Text and graphic intervention. On the other hand results indicated that embedded training interventions could help users to avoid phishing attacks. In the text and graphics condition, 80% of the users fell for the first phishing attack. Subsequently, 70% of the users clicked on the training email, and 70% of the users fell for the final phishing attack. But the mean percentage of users that fell for phishing over the last three phishing emails was 30% only.

The comic strip intervention was the most effective way in educating users to avoid phishing. On the downside, the comic strips were perceived as childish. 55% of participants preferred the text and graphics intervention above the comic strip (Kumaraguru, Rhee, Acquisti, et al., 2007). On the bright side, the comic strip was significantly more effective in teaching users phish avoidance behavior than the text and graphics intervention (Kumaraguru, Rhee, Acquisti, et al., 2007). All participants in the comic strip intervention condition fell for the first phishing email and clicked on the first training email. After training, only 30% of the users fell for the final phishing attack. Besides, the mean percentage of users that fell for phishing over the last three attacks was 23%.

-
- 1. Including an embedded design in anti-phishing training seems an effective way to enhance the ability of users to identify phishing emails.*
 - 2. Including a comic strip format in anti-phishing training seems to be a more effective way to enhance the ability of users to identify phishing emails than a text (and graphics) design.*
-

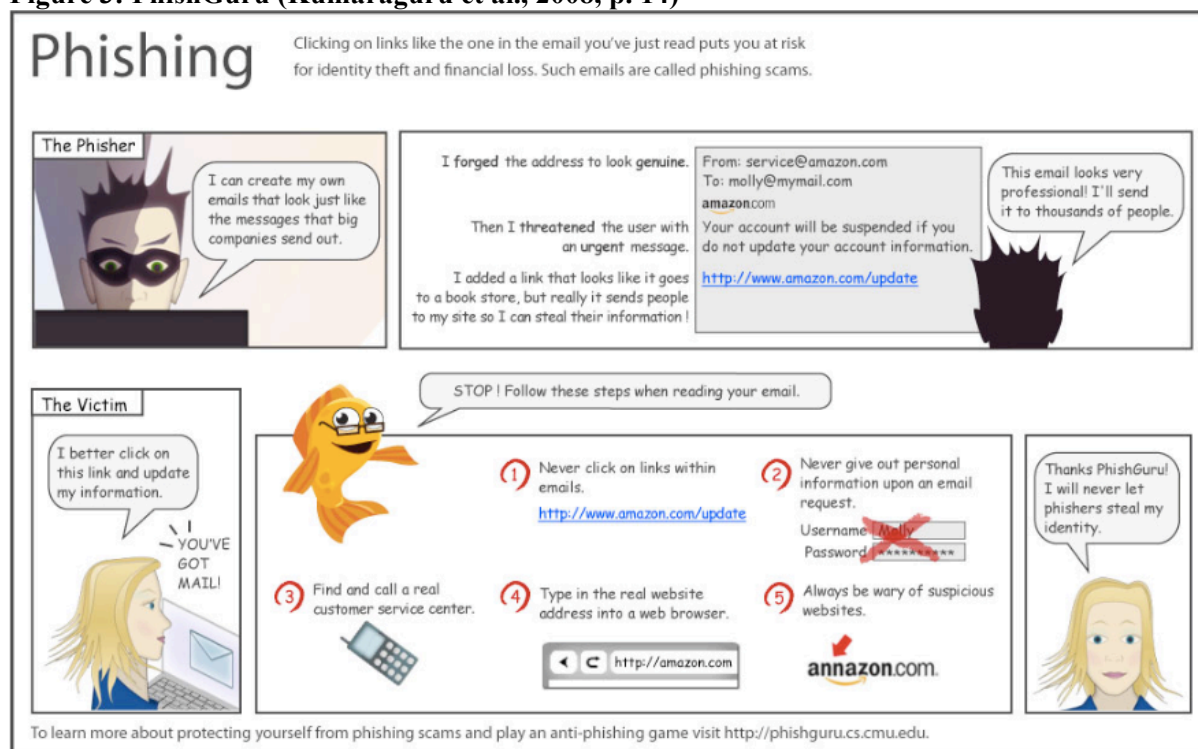
Too much text. That the comic strip outperformed the text and graphics intervention was explained as a result of the fact that the comic strip intervention used less text and more graphics (Kumaraguru, Rhee, Acquisti, et al., 2007). This may also explain the difference between two large-scale real-world corporate anti-phishing training studies that both examined the effect of embedded training (Caputo et al., 2014; Kumaraguru et al., 2008). One study used cartoon training (Kumaraguru et al., 2008) and the other used text training (Caputo et al., 2014). The cartoon training (using few words) increased phish avoidance behavior within the company. Conversely, the text training (using many words) did not prevent employees from falling for phishing.

Keep text in anti-phishing training simple and short seems an effective way to enhance the ability of users to identify phishing emails.

PhishGuru. The positive results led to further development of the comic strip intervention. The final version is called PhishGuru (Kumaraguru, Rhee, Sheng, et al., 2007). The content of PhishGuru is very similar to the earlier tested comic strip intervention. A few techniques to recognize phishing emails are combined with simple measures to prevent falling for phishing (figure 3).

The design of PhishGuru is a comic strip training intervention that uses avatars (a fish, a criminal, and a victim) to personalize the training. The fish helps the victim to escape from the criminal by giving tips, tricks and examples to avoid phishing emails (figure 3).

Figure 3: PhishGuru (Kumaraguru et al., 2008, p. 14)



Knowledge retention after one week. A second embedded training intervention study was performed with PhishGuru and included 42 students recruited around the Carnegie Mellon University. Like the previous study, students were given the role of Bobby Smith and had to process his email inbox (Kumaraguru, Rhee, Sheng, et al., 2007). Participants saw 16 emails before training, 16 emails in a direct posttest, and 16 emails in a delayed posttest after seven days (retention test). A retention test measured the ability to recall concepts learned in the past when tested under similar conditions after a period of time (Clark & Mayer, 2016). There were three treatment groups: a control group (did not receive training), a non-embedded group (saw phishing tutorial from Amazon), and an embedded group (saw PhishGuru) (Kumaraguru, Rhee, Sheng, et al., 2007).

Untrained users identified 7% of the emails correctly (legitimate or phishing) in the pretest, 11% in the direct posttest, and 7% in the delayed posttest (Kumaraguru, Rhee, Sheng, et al., 2007). Users in the non-embedded condition identified 4% of the emails correctly in the pretest, 14% in the direct posttest, 7% in the delayed posttest (Kumaraguru, Rhee, Sheng, et al., 2007). Users in the embedded training condition performed significantly better. These users identified 18% of the emails correct before training, 68% directly after training and 64% at the retention test (Kumaraguru, Rhee, Sheng, et al., 2007). These results support the conclusions of their previous study with the comic strip

intervention (Kumaraguru, Rhee, Acquisti, et al., 2007). Firstly, embedded training is an effective way to teach users about phishing. Secondly, embedded training is more effective than non-embedded training (Kumaraguru, Rhee, Sheng, et al., 2007). Thirdly, users trained by PhishGuru can retain their knowledge for seven days.

The results in these small-scale laboratory studies are questionable according to Parsons, McCormac, Pattinson, Butavicius, and Jerram (2015). They state that studies in which users are informed they take part in a phishing experiment are better able to distinguish legitimate emails from phishing emails. To deal with shortcomings of small laboratory studies, embedded training was tested in four larger real world studies (Caputo et al., 2014; Kumaraguru, Cranor, & Mather, 2009; Kumaraguru, Cranshaw, et al., 2009; Kumaraguru et al., 2008).

A third embedded training intervention experiment was again performed with PhishGuru but for the first time in a real-world corporate setting. Kumaraguru et al. (2008) used participants that worked at a large Portuguese company. The goal was to evaluate the effectiveness of anti-phishing training in a real-world corporate environment. All 321 participants of the study worked on the same floor of an office building. However, participants were from different areas in the firm: administration, business, design, editorial, management, technical, and others (Kumaraguru et al., 2008). To achieve their goal Kumaraguru et al. (2008) sent three simulated phishing emails to unknowing employees (at day 0, day 2, and day 7). The first email was to determine a base level of anti-phishing behavior and the following emails checked for improvement after training. If users clicked on the phishing email at day 0 they were provided with PhishGuru training according to the principles of embedded design. All emails were based on real phishing attacks that the company had received in the past (Kumaraguru et al., 2008). Fake phishing websites were linked to the phishing emails.

Kumaraguru et al. (2008) found that a significant amount of the users (42%) indeed clicked on links in phishing emails. Trained users (clicked on the first phishing email and were provided with PhishGuru training) were significantly (paired *t*-test, *p*-value <0.01) less likely to fall for the subsequent simulated phishing attacks. Only 19% of the trained users clicked and gave information during the second test and 12% of the users gave information during the retention test (Kumaraguru et al., 2008). These results showed that users did not significantly (paired *t*-test, *p*-value 0.55) lose any of their knowledge up to seven days in a real-world setting (Kumaraguru et al., 2008).

A control group existed out of employees that did not click on the link in the first phishing email and, therefore, did not receive training. 10% of the untrained users in the control group clicked and gave information in the second test, and 13% of the users clicked and gave information in the retention test (Kumaraguru et al., 2008). Kumaraguru et al. (2008) concluded that untrained employees were equally able in identifying phishing emails than trained employees, indicating that untrained employees did not need the training they had not received (Kumaraguru et al., 2008). This study, therefore, did not support previous laboratory research on anti-phishing training impact.

Spear training. Kumaraguru et al. (2008) also tested the effect of a spear version of the PhishGuru training. In this study, the spear phishing training differed from the generic training in a way that the spear training contained more detailed information (Kumaraguru et al., 2008). For example (Kumaraguru et al., 2008, p. 14): "*never give out personal information upon an email request*" (generic) or "*never give out corporate or financial information over email, no matter who appears to have sent it*" (spear).

As described before 42% of the users in the generic training condition clicked and gave information in the pretest, 19% of the users gave information in the direct posttest, and 12% during the retention test. On the other hand 39% of the users in the spear training condition clicked and gave information in the pretest and from these users, 18% gave information during the posttest one day later. Finally, 15% of the users clicked and gave information after one week. The result of the comparison showed no significant difference between the two conditions. The authors concluded that users did not gain specific abilities to identify spear phishing emails by being trained via spear interventions rather than generic interventions (Kumaraguru et al., 2008).

Focusing on spear phishing in anti-phishing training seems not an effective way to enhance the ability of users to identify phishing emails.

Knowledge retention after 28 days. The fourth embedded training intervention experiment also tested the effect of PhishGuru in a real-world setting. The experiment was performed with 515 active email users from the Carnegie Mellon University including student, faculty, and staff (Kumaraguru, Cranshaw, et al., 2009). These users were sent simulated phishing emails, not knowing they were taking part in a phishing experiment. The goal was to examine if people retain gained knowledge up to 28 days (Kumaraguru, Cranshaw, et al., 2009). All unknowing participants received three legitimate and seven simulated phishing emails in four weeks time. PhishGuru was again used to train users. There were three treatment groups: a control group (did not receive training), a one training condition (received PhishGuru training at day 0), and a multiple-training condition (received PhishGuru training at day 0 and day 14).

Kumaraguru, Cranshaw, et al. (2009) found that users who received PhishGuru training at day 0 only, performed significantly better than untrained users in avoiding the phishing attacks at day 28. 54.4% of the untrained users clicked on the link in the last phishing email, while only 27% of the trained users made that mistake. Therefore, the authors conclude that users could retain knowledge up to 28 days.

Repetitive training. Kumaraguru, Cranshaw, et al. (2009) also examined the difference between the single-training condition and the multiple-training condition. The results showed that an additional training message reduced the probability to fall for phishing attacks. 42.9% of the users in the one-training condition clicked on the link in the phishing email on day 16, while only 26.5% of

the users in the multiple-training condition fell for this attack. This significant difference remained until day 21. However, there was no significant difference between the single-training and multiple-training condition on day 28 (Kumaraguru, Cranshaw, et al., 2009).

Regardless of the design, content, or quality of training some studies provided evidence that retaining gained knowledge is difficult. Studies that tested retention of knowledge after 16 days (Alnajim & Munro, 2009b), four weeks (Lastdrager et al., 2017) or a few months (Canova et al., 2015; Caputo et al., 2014) presented insignificant results, indicating that knowledge fades away over time and, therefore, repetitive training is necessary.

Repetitive anti-phishing training seems necessary to enhance the ability of users to identify phishing emails over time.

Anti-phishing landing page. In a fifth experiment with embedded training interventions, PhishGuru was tested in a real-world setting as an anti-phishing landing page (Kumaraguru, Cranor, et al., 2009). This landing page was designed as a webpage to display on blacklisted websites. So after a phishing webpage was detected, it was removed from the Internet and replaced by the anti-phish landing page (Kumaraguru, Cranor, et al., 2009). The content and design of this training message was similar to PhishGuru (Kumaraguru, Cranor, et al., 2009). Despite the fact that the training was initiated and shown on websites, the training instructed users on how to identify phishing emails.

Monitoring the online behavior of 3,359 Internet users by tracking their IP-addresses, from January 2014 to April 2014, made it possible to measure the effect of the Anti-Phishing Landing Page (Gupta & Kumaraguru, 2014). Gupta and Kumaraguru (2014) compared the amount of times users clicked on blacklisted websites. They observed that clicking on blacklisted websites reduced by 46% in April as compared to January (Gupta & Kumaraguru, 2014). Therefore, Gupta and Kumaraguru (2014) conclude that the anti-phishing landing page was effective in educating users to avoid phishing attacks.

Two-column text training. Finally, a sixth experiment with embedded training interventions tested the effect of a two-column text training, and not PhishGuru, in a real-world corporate setting (Caputo et al., 2014).


The design of the two-column text training (figure 4) differed from PhishGuru on three main points: (1) the training used text only (no graphics), (2) the training used more text, and (3) the two-column text training did not include a storyline. The reason for text instead of a comic strip was that senior employees of the company felt that a comic strip intervention was not an appropriate format for corporate education (Caputo et al., 2014).

Despite the different design, the content of the two-column text training was very similar to that of PhishGuru. It explained why users were sent simulated phishing emails, what (spear) phishing was, and how users could avoid falling for phishing in the future. Again users were taught rigorous

measures as never click on links or attachments in emails. The two-column text training provided the following tips to identify phishing emails: "(1) mismatch between name and address, (2) motivation to take immediate action, (3) links do not match status bar, (4) improper grammar, odd spacing, and (5) the overall feeling that something is not right" (Caputo et al., 2014, p. 32).

Figure 4: Two-Column Text Training (Caputo et al., 2014, p. 5)

How to Defend against Spear Phishing



You have just been spear phished! The email that you just read was not actually from the **Wired** media alert list. It was a spear phishing email designed to help you learn how to protect your co-workers from cyber attackers.

How could you have recognized the spear phishing email you just received?

Spear phishing emails seem professional and legitimate. However, there are several ways to recognize them:

From: owner-media-alert-list@lists **Wired** org
on behalf of Rosetti, Mark C. <owner-media-alert-list@lists **Wired** org>

Sent: Tue 9/12/2011 12:00 PM

To: Doe, John

Subject: **Wired** makes "World's 50 Most Innovative Companies" list

Although we dropped to **Wired** in Fortune Magazine's "100 Best Companies to Work For" this year, we were just ranked #9 in Wired Magazine's "World's 50 Most Innovative Companies" list **and you'll never believe why**. Here is the link for those interested:

<http://www.wired.com/business/2011/07/innovativecompanies/>

I see **this a huge** feather in **Wired**'s cap.

Mark C. Rosetti
Lead Analyst/Manager/Engineer
Wired Corporation
mrosetti@wired.com (office)
mrosetti@wired.org

<http://www.wired.com/business/2011/07/innovativecompanies/>

- 1. What is spear phishing?**
Spear phishing is a form of cyber attack attempting to infiltrate your system or organization for cyber crime or espionage purposes. Such cyber attackers find inside information specifically relevant to you and craft fake email messages, usually impersonating well-known companies, trusted relationships, or contexts. In order for the attack to succeed, it requires that you take action. For example, by clicking on a link in the email message you could install malicious software on your system.
- 2. What do your co-workers stand to save when you don't fall for spear phishing attempts?**
By not clicking on links within spear phishing emails your co-workers save three things:
 - 1. Identity** - Your co-workers save their identity because cyber attackers can't access sensitive details (e.g., logins, passwords, etc.) from their systems.
 - 2. Time** - Your co-workers save their time because their systems won't have to be wiped and then restored with the last backup.
 - 3. Data** - Your co-workers save data because cyber attackers can't steal sensitive information from their systems.
- 3. What are simple ways to protect your co-workers?**
There are several easy things that you can do to protect your co-workers from spear phishing attacks:
 - **Never click on unanticipated links or attachments** within emails or forward/reply to emails asking for private information.
 - **Always verify contact information** by going directly to the source (i.e., using official phone numbers, emails, and websites instead of those provided)
 - **Report suspicious emails immediately** by calling the Help Desk, especially if you have clicked on the links provided.

*This research project is being conducted for a government sponsor and your identity will not be attached to any data results or be provided to **Wired** management. For more comprehensive **Wired** awareness material on spear phishing, visit [Wired](#).*

Thank you for your time and attention. Now that you have finished the training please close the browser.

Mismatch between name and address in "From:" field

Motivation to take immediate action

Links don't match status bar when mouse is hovered over

Typos, improper grammar, odd spacing

Intuition - overall feeling isn't right

To test their training 1,500 employees were randomly selected out of 6,000 employees from a medium-sized Washington, DC-based organization. Caputo et al. (2014) followed the methodology of Kumaraguru et al. (2008) where unknowing employees were sent simulated phishing emails to their corporate email accounts. In accordance with the embedded design, employees received training immediately after they clicked on links in simulated phishing emails. Two main methodological differences compared to the study performed by Kumaraguru, Cranshaw, et al. (2009) were: (1) there was no direct posttest, rather they performed retention tests only and (2) a tripled sample size was used. The goal of the study was to explore the effect of training in a corporate setting, while using a strong methodology (Caputo et al., 2014). Employees were sent three simulated phishing emails. The

first test was in February 2011, the second one in May 2011, and the third one in September later that year.

The results of this experiment did not support the findings of earlier studies on anti-phishing training impact. Firstly, in this study, the overall click rate was very high before training (Caputo et al., 2014). Where other studies showed a 30% click rate, the study of Caputo et al. (2014) showed an average click rate of more than 60% for the entire group. The difference may reflect the difficulty to recognize phishing elements in the spear phishing emails used by Caputo et al. (2014). Secondly, 11% of the users clicked on the phishing links in test one (before training), two and three (after training), regardless of their training condition. Also, approximately 22% of the users did not click on any links in test one (before training), two and three (after training) regardless of their training condition (Caputo et al., 2014). Thirdly, in contrast to previous studies, trained users did not perform significantly better than the control group (Caputo et al., 2014). The authors gave four possible explanations for the insignificant results. (1) Training has no effect in a corporate setting. (2) Repetition may be required to change behavior. (3) The presented training was ineffective. (4) Many users did not read the training material, and so it was hard to say if the training had an effect or not (Caputo et al., 2014). In line with the third option of having an ineffective training, according to the employees, the training was too dense with text, too cartoonish and had confusing colors (Caputo et al., 2014).

-
1. *Including an embedded design in anti-phishing training does not guarantee an enhanced ability of users to identify phishing emails.*
 2. *Include graphics in anti-phishing training seems to be an effective way to enhance the ability of users to identify phishing emails, but the use of too much text or confusing colors should be avoided.*
-

Another reason for the insignificant findings could be that the control group also received an embedded message after they clicked on a false link. The control group saw: "*You have just been spear phished. The email was not actually from... It was a spear-phishing email to raise your awareness regarding spear phishing emails.*" (Caputo et al., 2014, p. 32). As the act of sending this message to the control group may raise their awareness, it may also explains why both groups (train and control) increased performance but not statistically different from each other. 60% of the trained users and 62% of the users in the control group fell for phishing before training. After training only 34% of the trained users and 36% of the users in the control group fell for phishing.

Content of embedded training interventions. All the discussed embedded training interventions propose rigorous measures to avoid falling for phishing. Two examples are: never give out personal information upon an email request, and never click on links in emails (e.g. Caputo et al., 2014; Gupta & Kumaraguru, 2014; Kumaraguru et al., 2008). However, legitimate emails can also contain links and clicking on those links can bring convenience to users. Therefore, this avoidance

training is unwanted for most people who gain great benefits from conducting business online and do not want to handle all emails with such scrutiny (Downs et al., 2006). Teaching users to distinguish phishing emails from legitimate emails may be an alternative for the rigorous measures. If users know the difference between phishing emails and legitimate emails they can treat the emails accordingly. Hence, users will click on links in legitimate emails, but not click on links in phishing emails. Therefore, it is useful to know what characteristics phishing emails have.

1. The content of current embedded training interventions is limited in the amount of cues provided to identify phishing emails.

2. The content of current embedded training interventions provide some cues to identify phishing emails, but almost no cues to identify phishing URLs.

2.4 Game-Based Training

Another approach that is used to teach users to identify and avoid phishing attacks is by gaming. Gaming increases the motivation of users to learn (Sheng et al., 2007). In this review, only tested games are addressed, games that were in development but not tested are omitted (e.g. Hale, Gamble, & Gamble, 2015).

Anti-Phishing Phil. Phishing emails often contain links to phishing websites. However, most users are not aware of the structure of URLs and domain names (Herzberg & Jbara, 2008). Consequently, swindlers often succeed in tricking users to click on these links.

The most tested anti-phishing game is Anti-Phishing Phil. This game teaches users to distinguish between legitimate URLs and phishing URLs. The main message of this training is pay attention to URLs; they are good indicators of phishing. The main character in the Anti-Phishing Phil game is a fish named Phil. Phil eats worms that all contain specific URLs. Phil's job is to only eat the legitimate URLs and reject false URLs before running out of time (Sheng et al., 2007). Phil's father, in the meantime, gives tips to identify dangerous worms. The game exists out of four rounds, and every round starts with a short tutorial providing anti-phishing advice. Additionally, the training includes examples and practice questions (Sheng et al., 2007). Phil receives points when he eats legitimate worms and points are subtracted when Phil eats bad worms.

During this game users are taught the following things:

1. URLs with all numbers in the front are usually scam (Sheng et al., 2007).
 - [https:// 123.456.898.76 /ing/login](https://123.456.898.76/ing/login)
2. A URL has several parts (Sheng et al., 2007).
 - [https:// mijn.ing.nl/ internetbankieren/SesamLoginServlet](https://mijn.ing.nl/internetbankieren/SesamLoginServlet)
 - `https://` = Prefix
 - `mijn.ing.nl/` = Address
 - `internetbankieren/SesamLoginServlet` = File name
3. The most important part begins with `://` and ends with the next `/` (Sheng et al., 2007).

- <https://mijn.ing.nl/internetbankieren/SesamLoginServlet>
4. The address ends with the FIRST single /.
 - <https://mijn.ing.nl/internetbankieren/SesamLoginServlet>
 5. Within the address the right hand side is the most important (Sheng et al., 2007). It shows the site name.
 - <https://mijn.ing.nl/internetbankieren/SesamLoginServlet>
 6. A URL begins with http:// or https://, if the prefix contains an extra s, this indicates that the website is secure (Sheng et al., 2007). In addition a website is often considered as safe when it contains a security lock in the URL bar with matching certificate (Downs et al., 2006).
 - However, both the lock, as the "s" in "https://", do not guarantee safety according to Dong, Clark, and Jacob (2008) .
 7. When one is not a sure if a URL is phishing or legitimate, one can always use Google to search for the site name (not the entire addresses). The first hit in Google should be a legitimate website (Sheng et al., 2007).

The Anti-Phishing Phil game was tested in five studies (Arachchilage, Love, & Beznosov, 2016; Davinson & Sillence, 2010; Kumaraguru et al., 2010; Mayhorn & Nyeste, 2012; Sheng et al., 2007). Sheng et al. (2007) were the first to test the effect of Anti-Phishing Phil. 42 participants were recruited via flyers around the Carnegie Mellon University campus, with recruitment emails, on university bulletin boards, and via craigslist.com (Sheng et al., 2007). Users were presented 10 websites and asked if the websites were legitimate or phishing and how confident they were in their judgment (scale 1 to 5). After the first test, users played Anti-Phishing Phil for 15 minutes and then they were asked again to judge 10 websites (Sheng et al., 2007). Results were compared with users in an existing training material group that saw anti-phishing tutorials from eBay, Microsoft, and the MySecureCyberspace portal (Sheng et al., 2007).

Anti-phishing tutorials. The study by Sheng et al. (2007) showed that users in the existing training materials group identified 66% of all websites correctly (legitimate or phishing) in the pretest and 74% during the posttest. Users, who played Anti-Phishing Phil, identified 69% of all websites correctly before training and 87% after training, which was a statistically significant improvement. A comparison between the two treatment groups showed that the game condition performed significantly better in the posttest than the existing training material group (Sheng et al., 2007).

Including interactivity in anti-phishing training seems to be a more effective way to enhance the ability of users to identify phishing URLs than the use of passive anti-phishing tutorials.

The effect of Anti-Phishing Phil was later examined in a very similar experiment (Kumaraguru et al., 2010). However, this study included 4,517 participants. The results of a control group (no training, 2,496 people) were compared with a trained group (Anti-Phishing Phil, 2021

people). Participants in the game condition saw six websites before playing Anti-Phishing Phil (pretest), six websites after playing the game (direct posttest), and six websites one week later (delayed posttest). So in total they saw 18 websites, in three phases, with three legitimate websites and three phishing websites (Kumaraguru et al., 2010).

The results showed that users who scored poorly in the pretest (maximum four websites were estimated correctly) improved their ability to identify phishing websites significantly. Before training users identified 57% of the phishing websites incorrectly as legitimate, while after training users identified 22% of the phishing websites incorrectly as legitimate (Kumaraguru et al., 2010). So users made less crucial mistakes (identifying a phishing websites as legitimate websites) after training than before training. However, the study does not mention if the control condition improved in the posttest as compared to the pretest.

The success in these experiments led to the development of a version of the Anti-Phishing Phil game for the smartphone (Arachchilage & Cole, 2011; Arachchilage & Love, 2013). Like the computer version, the smartphone version showed in a test that it enhanced the ability of users to identify phishing emails (Arachchilage et al., 2016). Before playing the game, participants identified 56% out of 20 websites (phishing or legitimate) correctly, while after playing the game they identified 80% out of 20 websites correctly.

Teaching users the structure of a URL and typical characteristics of phishing URLs seems to enhance the ability to users to identify phishing URLs.

Two studies found less promising results with Anti-Phishing Phil (Davinson & Sillence, 2010; Mayhorn & Nyeste, 2012). Davinson and Sillence (2010) investigated the effect of education by Anti-Phishing Phil in combination with risk level manipulation on day-to-day online behavior. By means of questionnaires 64 psychology students were asked how secure they behaved online in their private lives.

Risk level manipulation and game-based training. The experimental design was as follows. First, in a baseline measurement users were asked how secure they acted online in the past week. Then students were given a warning message that was either: *"Well done! You have a low percentage risk of becoming a victim of fraud due to the way you use the Internet."* or *"Warning! You have a very high percentage risk of becoming a victim of fraud due to the way you use the Internet."* (Davinson & Sillence, 2010, p. 1743). Subsequently students were asked in an intention measure how secure they intended to behave online the coming week (Davinson & Sillence, 2010). After this intention measure, half of the participants received anti-phishing training by Anti-Phishing Phil. One-week later students were again asked how secure they had behaved online in the past week.

On the one hand, the results showed that when users received a risk warning, the intention was to behave more secure. Besides, the follow-up test after a week confirmed that students acted

significantly safer as compared to the baseline measurement (Davinson & Sillence, 2010). On the other hand, the retention test confirmed that students did not act as careful as intended. Moreover, the improvement of secure behavior for students with training was not statistically different from the students that did not receive training. Therefore, the paper concluded that playing the Anti-Phishing Phil game did not have a significantly higher impact on protecting users from phishing than a simple risk message.

Two possible reasons for the insignificant finding are: (1) Davinson and Sillence (2010) did not actually monitor behavior, but merely asked if students acted more secure. (2) All users received a risk warning that may have enhanced their fear for phishing making them more cautious online (Davinson & Sillence, 2010).

Fear for phishing. Zielinska et al. (2014) tested the idea that increased fear for phishing, rather than training, would result in phish avoidance behavior. Zielinska et al. (2014) performed an experiment with 96 participants, recruited via Amazon Mechanical Turk. The participants had to identify eight emails before training and 12 emails after training. Three training conditions were compared. All conditions received the same anti-phishing information, but the three groups differed on the extra information they received. The control group watched a cooking video; the other two groups received information that aimed to increase fear. One group saw a video on real-world consequences of phishing to increase specific fear for phishing. The other group saw news articles to increase fear in general (Zielinska et al., 2014).

The results indicated that the level of fear for phishing had no impact on phish avoidance behavior. Hence, there was no significant difference between the treatment conditions (Zielinska et al., 2014). However, participants in all conditions showed some improvement (correctly identified emails / total amount of emails) in the second test (ratio's around 0.6) as compared to the first test (ratio's around 0.55) indicating that training, rather than fear, enhanced secure behavior.

Stimulating fear in anti-phishing training does not seem to be an effective way to enhance the ability of users to identify phishing emails.

Interventions and game-based training. The last published experiment performed with Anti-Phishing Phil included 84 psychology students from the North Carolina State University (Mayhorn & Nyeste, 2012). In this study, participants in the game condition ($n = 28$) played Anti-Phishing Phil before they had to interact with an email inbox that contained 30 emails. The participants in the game condition also saw embedded training materials, in the form of an undefined cartoon, when they clicked on links in phishing emails (Mayhorn & Nyeste, 2012). After one-week, a retention test was performed with 40 emails (from which 10 emails were new).

The study does not provide information on mean scores for the various training conditions. Nevertheless, the study concludes that users that received training performed significantly better than

the control group that did not receive training (repeated measures ANOVA with an alpha level of 0.05). The positive effect of the training remained in the second test, however, this time not statistically significant different from the control group (Mayhorn & Nyeste, 2012), possibly because of the small population.

In sum, two experiments did not find a main effect of Anti-Phishing Phil training (Davinson & Sillence, 2010; Mayhorn & Nyeste, 2012). In both studies users behaved more secure after anti-phishing training than they did before, yet these positive results were also found for the control group.

NoPhish. Canova et al. (2015) developed a different kind of game in the form of an application for the smartphone. The game existed out of questions on URLs. Users of this game had three lives represented by hearts (Canova et al., 2015). Points were gained when questions were answered correctly and lost when answered wrongly.

The content of NoPhish was similar to Anti-Phishing Phil. In eight lessons users were taught the structure of URLs, obvious subdomain tricks (ing.phishing.nl), IP address tricks, random subdomain tricks (ing.monypomy.nl), trustworthy sounding domain tricks (ing.secure-login.nl), typos in domain names (twitter.com), substitute characters in domain names (amazon.com), and brand name in path tricks (abs.nl/ing) (Canova et al., 2015).

An experiment was performed to test the game. Participants, recruited via flyers and social media, had to identify 16 websites before training and 24 websites after training (eight new websites) as phishing or legitimate. In between rounds, users played the NoPhish game for 30 minutes (Canova et al., 2015), which was longer than in other studies where users were trained for 15 minutes (Sheng et al., 2007).

The study showed that users performed significantly better after training (90% correct) than before training (57% correct) in identifying the websites (Canova et al., 2015). The study also demonstrated that after playing the game, participants would use URLs as their main indicator to judge about a website's legitimacy (Canova et al., 2015). A retention test measured if users could retain knowledge up to five months. On the one hand, performance was still significantly better (81% correct) than before participants played NoPhish. On the other hand, participants significantly had fewer correct answers compared to the direct posttest (Canova et al., 2015).

Conversely, a study that compared classroom training, text based training, and computer-based training (using NoPhish) showed less positive results (Stockhardt et al., 2016). Participants trained by NoPhish performed better after training (81.5% correct) than before training (57% correct) in identifying phishing websites. However, the performance after training was statistically less than the performance of participants that followed classroom training who increased their performance from 65% to 94% correctness (Stockhardt et al., 2016).

Anti-Phishing Education Game. Another anti-phishing game is called Anti-Phishing Education Game (APEG) (Yang et al., 2012). The main character in this game was a soldier named John. John had to determine if various hyperlinks were phishing or not (Yang et al., 2012). John's

commander provided tips on how to identify the legitimacy of a website. The main difference with Anti-Phishing Phil was that this game used a realistic look (the user of the game saw multiple hyperlinks after a Google search). The rationale for the realistic layout was that it was needed to be able to apply gained knowledge in the real world (Yang et al., 2012).

Because anti-phishing knowledge is increasing Yang et al. (2012) classified the anti-phishing knowledge into different levels of learning skills. For each level they proposed a matching action. Firstly, users were educated to be aware of the risk of domain names. Users could identify the nature (legitimate/phishing) of a domain name by checking black and white lists. Black lists reveal phishing websites and can be found via Phish Tank. White lists reveal legitimate websites and can be found via a Google search (Yang et al., 2012). Secondly, users are educated to pay close attention to ads and login fields on websites. These fields could be spoofed. Thirdly, users had to learn how to detect false URLs from legitimate URLs by decoding them (Yang et al., 2012). Finally, Yang et al. (2012) warned users for domain name spoofing by replacing characters, for example by using zero "0" instead of alphabet "o".

A pilot test in which 62 students from a large Taiwanese University had to judge about 20 websites before training and 20 websites after training showed positive results. The students were significantly less likely to fall for phishing after training than they were before training (paired t -test < 0.001). However, the control group (no training) also improved their phish avoidance behavior (paired t -test = 0.001) (Yang et al., 2012). According to Yang et al. (2012) this indicates that the posttest was easier than the pretest.

Other anti-phishing games. Two other anti-phishing games that used quizzes as their main format, have been developed and tested. Smith et al. (2009) proposed an anti-phishing website (Social-Ed) and Sercombe and Papadaki (2012) an anti-phishing game called Malware Man.

The Social-Ed website was tested in a pilot study including 46 students of the University of Sidney. The results showed that participants that read the reading material on the Social-Ed website performed better in answering phish related quizzes (69% correct) than users that did not read the training materials (44% correct) (Smith et al., 2009).

The Malware Man game was tested with 104 students and staff members from the Plymouth University (Sercombe & Papadaki, 2012). In this game, users had to answer phish related questions. The game used a graphic displaying a malware man behind a low firewall. As the users answered more questions correctly, the firewall grew, causing the malware man to be in pain (Sercombe & Papadaki, 2012). After playing the game, users had to complete a survey with phish related questions. The answers were compared with a control group that did not receive any training. The results indicated that users who played the game answered 77% of the posttest survey questions correctly. While users that did not play the game answered 55% of the questions correctly (Sercombe & Papadaki, 2012).

Using a game format in anti-phishing training seems to be an effective way to enhance the ability of users to identify phishing URLs.

Content of anti phishing games. Most anti-phishing games focus on recognizing phishing URLs. Games as NoPhish and Anti-Phishing Phil teach users the structure of URLs and how this structure differs for legitimate URLs as compared to phishing URLs.

Phishing URLs are also examined by studies that developed technical countermeasures for phishing. Most of these technical countermeasures implement techniques that are unpractical or difficult to teach users like time taking activities as counting the number of slashes, dots, or characters (e.g. Aggarwal, Rajadesingan, & Kumaraguru, 2012; Jeeva & Rajsingh, 2016), identifying the age of domain names (Basnet, Mukkamala, & Sung, 2008), or combining many indicators of phishing to form a legitimate judgment (Jeeva & Rajsingh, 2016). Nevertheless, technical countermeasures confirm that characteristics of phishing URLs taught in current anti-phishing games, as Anti-Phishing Phil and NoPhish, are useful. These technical countermeasures confirm that URLs with the following characteristics are usually phishing:

- URLs that use an @. The use of "@" leads the browser to ignore everything proceeding the @ (Ahmed & Abdullah, 2016).
- URLs that use an IP address (Ahmed & Abdullah, 2016; Jeeva & Rajsingh, 2016).
- URLs that use a hyphen (-) in the website address (Ahmed & Abdullah, 2016; Jeeva & Rajsingh, 2016).

Phishing URLs often use an @, an IP address, or a hyphen (-) in the website address.

Therefore, the content of current anti-phishing training materials seems correct and helpful in teaching users to distinguish phishing URLs from legitimate URLs. However, the game-based training programs do not consider other forms of phishing. Hence, users are not taught about phishing emails in general, without URLs.

The content of current anti-phishing games is limited in the amount of cues on how to identify phishing emails (without links).

2.5 Most Effective Training Techniques

The literature review indicates that the two most tested training techniques are (1) sending simulated phishing emails in combination with an error message, warning message, or training intervention, and (2) game-based training. Embedded training interventions are most suitable to train inexperienced email users or people with high web reliance because these interventions increase awareness (Abbasi, Mariam Zahedi, & Chen, 2016). Game-based training is made to provide more advanced anti-phishing knowledge, which is useful to train aware users that may be overconfident in

their ability to avoid phishing attacks (Abbasi et al., 2016). To combine embedded training and game-based training, an anti-phishing game could, for example, be initiated after users falls for a simulated phishing attack. Some games, that combine the benefits of embedded training and game-based training, have been developed but not yet tested (Hale et al., 2015). These games may prove to be useful in the future.

Error message, warning message, or embedded training intervention. Studies showed that sending simulated phishing emails followed by an error message, a warning message or a training intervention can enhance phish avoidance behavior.

To identify the most effective way to train users Dodge et al. (2012) performed a test with 892 students from the United States Military Academy. They tested the difference between a control group, and two experimental groups who (1) received feedback or (2) were trained. All students were sent three simulated phishing emails with links to websites (Dodge et al., 2012). There was a baseline measurement, a 10-day follow-up measurement, and a measurement after 63 days. The control group contained 287 students. If students in the control group entered data into a website they saw an error message. The experimental group that received a notification contained 298 students. These students received feedback on how they could have identified the email as phishing. Finally the experimental group that received training contained 307 students. These users were sent a phishing awareness training if they entered data into the website (Dodge et al., 2012).

The results showed that there was no significant difference in phishing susceptibility between any of the treatment groups after 10 days (Dodge et al., 2012). Overall the 10-day retention test showed a very low percentage of victims (approximately 9% of the students in each condition), indicating that the email was not tempting for students to click on (Dodge et al., 2012). However, after 63 days, the treatment groups performed significantly different. The training condition performed best and 75.5% of the students in this group avoided the third phishing attempt. This percentage was 67.92% for students in the notification group and 52.5% for students that saw an error message. Nonetheless, all conditions improved as compared to the baseline measurement (Dodge et al., 2012). So, while feedback led to increased awareness, embedded training was the most effective way to enhance secure behavior (Dodge et al., 2012). Moreover, Alnajim and Munro (2009b) tested the effect of a warning message and did not find significant retention of knowledge after 16 days. While Kumaraguru, Cranshaw, et al. (2009) tested the effect of an embedded training intervention and did find significant retention of knowledge after 28 days.

An embedded training intervention seems to have a greater effect on the ability of users to identify phishing emails than an error message or a warning message.

As was mentioned previously, Caputo et al. (2014) showed in their study that a two-column text training did not perform better than a control group who saw a warning message. Therefore, the

success of a training intervention depends on its design. The most successful intervention is PhishGuru, which showed in various experimental designs that it could enhance phish avoidance behavior (Kumaraguru, Cranshaw, et al., 2009; Kumaraguru et al., 2010).

Game-based training or (embedded) training intervention. Sheng et al. (2007) performed a study that examined if a game design would be more effective than printed training materials. In this study, results of a game-condition (played the Anti-Phishing Phil game) were compared with a tutorial condition (saw printouts of the lessons in Anti-Phishing Phil) and an existing training material condition (read phishing tutorials from eBay and Microsoft). The results showed that before training there was no significant difference between the groups in the total amount of correctly identified websites. After training, the game condition performed the best and significantly better than the existing training material condition (Sheng et al., 2007). The tutorial condition did not perform statistically significant different from the existing training material condition (Sheng et al., 2007). The study does not mention if there is a significant difference between the tutorial condition and the game condition in total correctly identified websites. Besides, the population was very small ($n = 14$ per condition) and the study did not find any statistical differences between the tutorial condition and the game-condition in making the dangerous mistake of identifying phishing emails as legitimate emails (Sheng et al., 2007).

Another study tested the difference between participants that received embedded training in the form of a cartoon and participants that received embedded training in the form of a cartoon in combination with game-based training (in the form of Anti-Phishing Phil) in identifying phishing emails (Mayhorn & Nyeste, 2012). In this laboratory setting participants had to process an email inbox as they would normally do (Mayhorn & Nyeste, 2012). The experiment was unable to identify a statistical difference between the two treatment groups in the direct posttest as well as in the retention test one week later (Mayhorn & Nyeste, 2012).

In sum, it remains difficult to determine the exact effect of anti-phishing games in comparison to (embedded) training interventions because many of the developed anti-phishing games were tested in small-scale pilot studies (e.g. Sheng et al., 2007; Yang et al., 2012). On the bright side, most anti-phishing experiment with games show positive results in teaching users to identify phishing attacks (appendix 1 and appendix 2). Besides, the positive effect of learning by gaming is confirmed in learning science (Clark & Mayer, 2016).

2.6 Other Proposed Anti-Phishing Training Methods

Several studies experimented with other methods to provide anti-phishing training. These studies suggest tablets and worksheets (Sun & Lee, 2016), classroom training (Stockhardt et al., 2016), posters (Kritzinger, 2016), or visualization tools (Zhang-Kennedy, Fares, Chiasson, & Biddle, 2016) as human oriented countermeasures for phishing.

Potentially effective anti-phishing training methods. Sun and Lee (2016) examined the effect of tablets and worksheets on motivation to learn about phishing including 155 eight- and ninth-

grade students. All students performed a pretest phishing questionnaire that measured their learning motivation and learning performance. Then, the children received 55 minutes of anti-phishing training in a classroom setting combined with 20 minutes of worksheet training (control group) or 20 minutes of concept map training (experimental group). The experiment showed that when pretest anti-phishing knowledge scores were low, users in the experimental condition obtained statistically significant higher posttest scores than users in the control condition ($p < 0.05$). However, when the pretest scores were high there was no statistically significant difference between the two treatment groups in posttest scores (Sun & Lee, 2016).

Posters (Kritzinger, 2016), or classroom training (Stockhardt et al., 2016; Lastdrager et al., 2017) are also proposed as a solution to phishing. Robila and Ragucci (2006) found a positive effect of class discussions on scores in phishing IQ tests. In phishing IQ tests users have to distinguish fraudulent emails from legitimate emails (Robila & Ragucci, 2006). Lastdrager et al. (2017) examined the effect of classroom training on the ability of children to discriminate phishing emails and websites from legitimate emails and websites. They found, after a 40-minute anti-phishing presentation and discussion session, that trained children improved their ability to correctly discriminate emails as legitimate or phishing by 14% (Lastdrager et al., 2017). However, after four weeks the enhanced phishing detection ability of trained children returned to pre-training levels (Lastdrager et al., 2017). Moreover, for companies classroom training may be time-consuming and expensive. Besides, computers offer opportunities for unique engagement with simulations of contexts that are not possible to replicate outside the digital environment (Clark & Mayer, 2016).

Finally, an interactive information visualization-tool called Geo-Phisher was developed and tested with 30 university students (Zhang-Kennedy et al., 2016). The information visualization tool presented data and information about phishing in a visual form (Zhang-Kennedy et al., 2016). The goal of this visualization-tool was to spark curiosity in data on phishing and to get the general public acquainted with the problems of phishing (Zhang-Kennedy et al., 2016). The tool also provided information on how phishing works and some tips to avoid becoming a victim (Zhang-Kennedy et al., 2016). The effect of Geo-Phisher was tested with two questions asked before and after training: (1) "*what is phishing?*" and (2) "*can you describe what you know about how to protect yourself from phishing?*" (Zhang-Kennedy et al., 2016, p. 9). Question 1 was answered correctly 80% before users saw the content of Geo-Phisher and 88% afterwards ($N = 15$). Question 2 was answered correctly by one participant before training and by 11 participants after training (Zhang-Kennedy et al., 2016).

Adverse effect of training. Anti-phishing training does not always work. Kearney and Kruger (2013; 2014) did two similar studies in which unknowing employees were sent a simulated phishing emails to their corporate email account. The emails requested for a username and password (with slightly different contexts). There were 1,700 active email users during the first experiment (Kearney & Kruger, 2013) and 1,400 active email users during the second experiment (Kearney &

Kruger, 2014). The studies measured how many of the users, that fell for phishing, had followed corporate anti-phishing training in the past.

In the first study 69%, of the users who gave their username and password had followed corporate anti-phishing training (Kearney & Kruger, 2013). In the second study the percentage was 92% (Kearney & Kruger, 2014). These results indicated an adverse effect where trained users fell for phishing more often than untrained users (Kearney & Kruger, 2014). Kearney and Kruger (2014) explained the result by pointing out that users trusted and relied on the security firewall of their company. Confidence in the company's protection lowered the employees' cautiousness towards phishing (Kearney & Kruger, 2013; 2014). Unfortunately, Kearney and Kruger (2013; 2014) did not train employees, but simply asked if employees had followed corporate training.

2.7 Overview of Findings

In conclusion of the literature review, many anti-phishing experiments have been performed. Table 1 (p. 10) gives an overview of the most important developments and findings within anti-phishing training research. A more detailed overview, that also contains information on experimental setups, populations, and actual results (in percentages), can be found in appendix 1 and appendix 2.

Most studies show that increased knowledge as a result of anti-phishing training will enhance the ability of users to identify and avoid phishing attacks. In general current digital anti-phishing training experiments can be classified in two categories: (1) sending simulated phishing emails in combination with an error message, warning message or training intervention to teach users about phishing emails, and (2) game-based training that teach users about phishing URLs.

The success of the most tested embedded training intervention PhishGuru can be explained by their comic strip design. Other embedded training interventions that used fewer graphic, more text, or included more details were less successful. When considering the content of current anti-phishing interventions two things stand out: (1) current interventions focus on tips and tricks to recognize phishing emails, but not phishing websites. (2) The tips and tricks to avoid phishing emails are rigorous (never click on links in emails, or never give personal information).

The most tested anti-phishing game is Anti-Phishing Phil, it showed in multiple experiments that it enhanced the ability of users to distinguish between phishing URLs and legitimate URLs. Other anti-phishing games that showed positive results during pilot tests used similar designs as Anti-Phishing Phil. These games used a combination of lessons, practice questions and feedback moments. When looking at the content of anti-phishing games they do not consider phishing emails. Both embedded training interventions and game based training were successful, yet a combination of these techniques has not yet been tested. Hence, an embedded game-based training that teaches users about phishing emails and phishing URLs may prove to combine the best of both techniques.

In sum, according to current anti-phishing studies, an effective anti-phishing training with the purpose to enhance the ability of users to identify and avoid phishing emails has an interactive, repetitive, embedded design that does not use much text or details, but rather uses a comic strip

format to educate users. The content of this training contains cues to identify phishing emails and cues to identify phishing URLs as well as a solution for uncertain situations (Table 2).

Table 2: Characteristics Central to the Effectiveness of Anti-Phishing Training

	Characteristics central to the effectiveness of anti-phishing training							
	Content			Design				
	Cues in emails	Cues in URLs	Emergency solution	Interactive	Embedded	Repetition	Cartoon	Simple and short
Tested training interventions	√		√		√	√	√	√
Tested training games		√	√	√			√	√
Ideal training	<u>√</u>	<u>√</u>	<u>√</u>	<u>√</u>	√	√	<u>√</u>	<u>√</u>

Note: the underlined characteristics central to the effectiveness of anti-phishing training in the ideal training row (√) are implemented in the anti-phishing training that was developed for this thesis.

3. DEVELOPMENT AND TEST OF NEW TRAINING MATERIAL

Based on conditions of effective training this chapter describes the development of PHREE, a new anti-phishing training, and how it was tested. The goal of this pilot test was to examine if the developed anti-phishing training could enhance the ability of users to identify phishing emails.

3.1 Development of Anti-Phishing Training PHREE

This section presents the objective of PHREE as well as its content and design features, and how it addresses learning science principles.

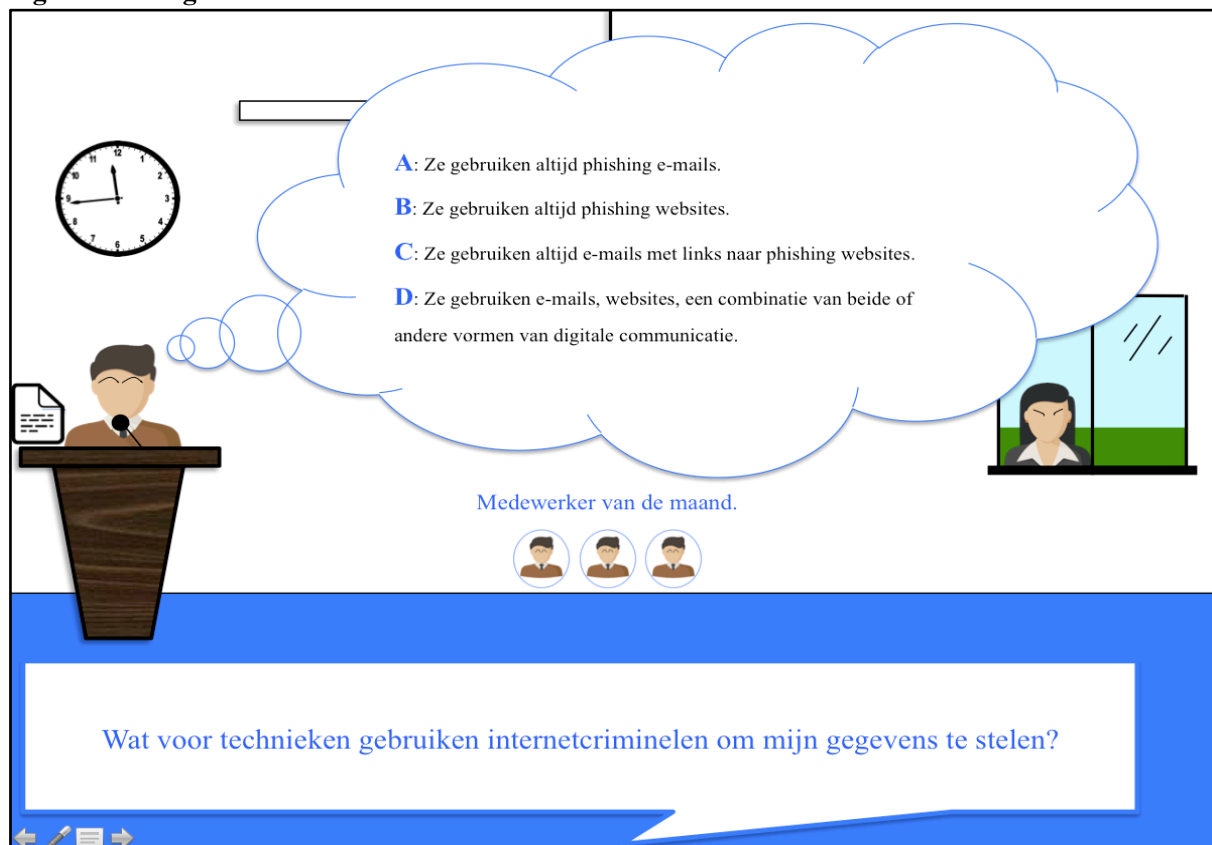
The objective of PHREE. The goal of PHREE is to teach users (1) cues of phishing emails and (2) cues of phishing URLs to increase the ability of users to identify phishing emails. As mentioned in section 2.7, some studies focused on identifying phishing emails, other studies concentrated on phishing URLs. Yet, studies that combined email and URL training are scarce (appendix 3). Zielinska et al. (2014) tested a training that included cues on emails and URLs, but this study examined the impact of fear for the consequences of phishing, not training. Mayhorn and Nyeste (2012) performed a study in which one experimental condition received email training and URL training. However, they made use of two training techniques (Anti-Phishing Phil and an undefined comic strip intervention). Since most phishing emails contain links to phishing websites, the goal of PHREE is to teach users cues of phishing emails and cues of phishing URLs to identify phishing emails. The game is made for anyone that wants to be able to distinguish phishing emails from legitimate emails independently. Therefore, the training is suitable for individuals who find it

hard to assess the reliability of emails, but can also function as business training to increase the level of security of an organization.

Characteristics of effective training in PHREE. To achieve this goal PHREE implements characteristics of effective anti-phishing training as stated in table 2 as well as learning science principles.

Design of PHREE. The design of PHREE is a game based training developed in PowerPoint (figure 5). A game format is chosen because multiple experiments showed it is an effective way to educate users (e.g. Arachchilage et al., 2016; Canova et al., 2015; Sheng et al., 2007), and a game format is supported by learning science (Clark & Mayer, 2016). Besides, it gives the possibility to teach users more detailed information than a training intervention. In line with characteristics of effective training PHREE is interactive, kept simple and short, and has a cartoon-based design. An embedded design and repetition are recommended, but not implement due to the limited time frame to perform this research.

Figure 5: Design of PHREE



Content of PHREE (Feedback). To determine what content anti-phishing training should have, studies examined phishing cues in emails and URLs (e.g. Kumaraguru, Rhee, Acquisti, et al., 2007; Marett & Wright, 2009). Most of the phishing emails contain a request for personal information, either directly or via a link to phishing websites in the email (Downs et al., 2006).

Phishers use all kinds of deceptive techniques to reveal this request. These deceptive techniques are addressed in four rounds of training.

- Lesson 1: phishing and how to avoid it at all time. In this lesson users are presented an example of how phishing works. Users are given an example of a criminal who sends a phishing email to potential victims and how easy it is for a potential victim to fall for such attacks. Finally, in this round users are taught what to do, when they are insecure about the legitimacy of an email. The advice is: look up a company in a trusted source like the Yellow Pages and call customer service.
- Lesson 2: characteristics of phishing emails. In this round, users are taught cues of phishing emails. Users are taught to pay attention to (1) the sender of an email, (2) the content of an email (request for personal info, need for a quick response, threats for not replying or rewards for replying, and typos), and (3) links in emails.
- Lesson 3: structure of a URL. In this round users are taught what the most important part of a URL is and how to recognize phishing domain name tricks.
- Lesson 4: characteristics of phishing URLs. In the last round of training users are taught about domain name spoofing (replacing the alphabet 'i' by number '1', or alphabet o by number 0).

Each round of training consists out of an instructional video combined with four related multiple-choice (A, B, C, or D) practice questions. Users receive feedback directly after each question explaining in detail why each multiple-choice answer was (in)correct.

The main difference between PHREE and existing training interventions is that the design of PHREE is interactive and the content of PHREE contains cues to identify phishing emails and phishing URLs. Current training interventions (e.g. PhishGuru) are passive training materials that focus on cues to identify phishing emails only (table 2). The main difference between PHREE and existing anti-phishing games is that the content of PHREE contains cues to identify phishing emails and phishing URLs. Current anti-phishing games (e.g. Anti-Phishing Phil) focus on cues to identify phishing URLs (table 2).

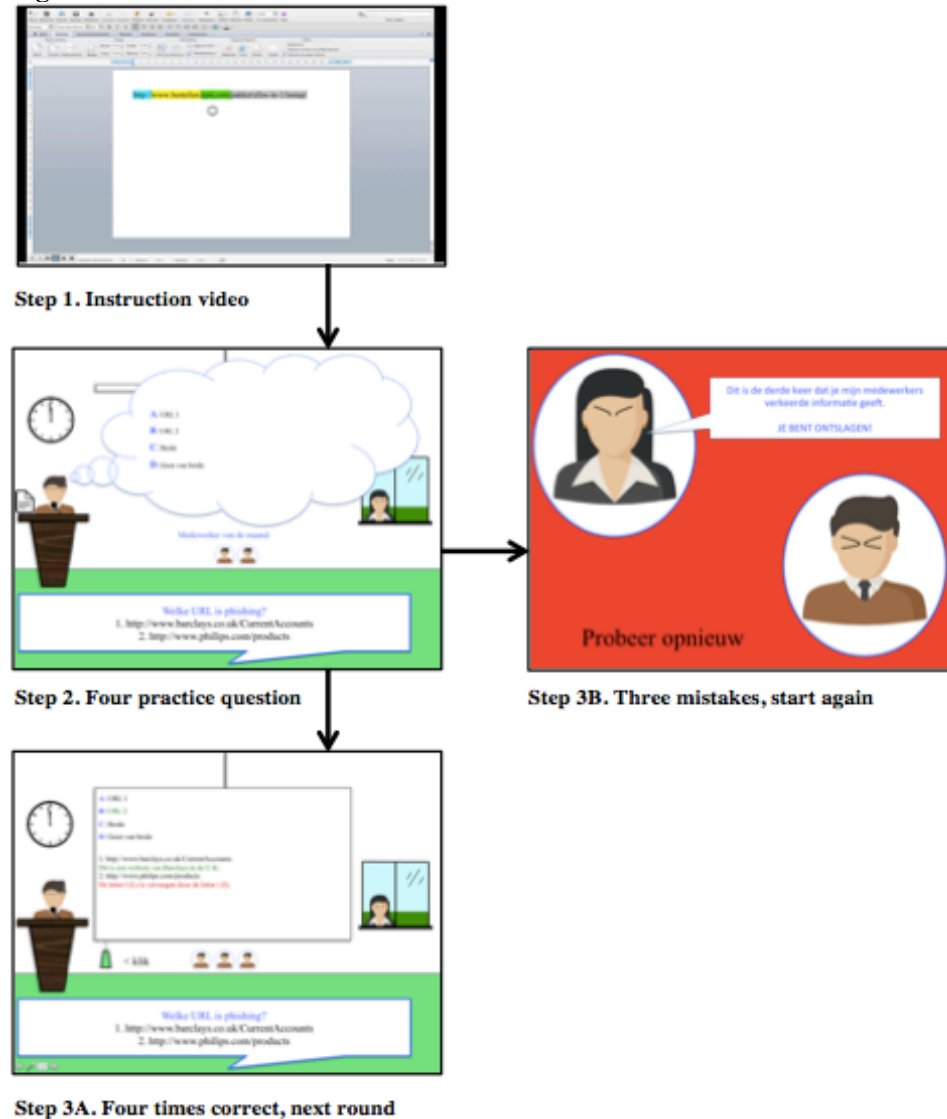
Learning Science Principles in PHREE (Flow). PHREE also addresses several learning science principles. Learning science suggests the following principles for an effective online training (Clark & Mayer, 2016):

- According to the Multimedia principle, training has to combine words and graphics and use instructional graphics rather than decorative graphics (Clark & Mayer, 2016). Therefore, users are presented instructional videos in PHREE that contain instructions on how to identify cues in phishing emails and phishing URLs, clarified by real phishing examples.
- The contiguity principle states that guiding texts should be placed close and next to graphics. The principle also states that training questions and feedback on those questions should be presented on the same page (Clark & Mayer, 2016). In PHREE, practice questions are asked after each instruction video (figure 5). If users answer these questions wrongly they are

presented with immediate feedback.

- According to the modality principle, words should be presented in audio rather than by onscreen text to describe a graphic (Clark & Mayer, 2016). Therefore, the instructional videos in PHREE are supported by audio narration.
- The Redundancy principle states that visuals should be explained with words or by text, but do not use both (Clark & Mayer, 2016). Therefore, PHREE does not use written text to support the instructional videos.
- In line with the coherence principle game-based training should avoid any information that does not support the instructional goal, for example background music (Clark & Mayer, 2016). This principle is implemented in a way that there is no background music in PHREE and the design is kept as basic as possible.
- The personalization principle states that training should use of a conventional style and virtual coaches in the form of agents to personalize training (Clark & Mayer, 2016). The PHREE training uses words as 'I', 'you', 'yours', has a cartoon format and includes an agent.
- The segmenting- and pre-training principle indicates that training should be split-up in parts. Segmenting is breaking a lesson into parts that are manageable rather than one continuous element. Pre-training are instructions before the actual training so that the names and characteristics of the main concepts are clarified (Clark & Mayer, 2016). PHREE meets these requirements and starts by introducing the characters, then provides game instructions, and finally presents training materials segmented over four rounds.
- A game-based training should include worked examples. Worked examples are examples that show step-by-step how to solve a problem or how to perform tasks (Clark & Mayer, 2016). The instructional videos in PHREE present worked examples on how phishing works, how to recognize phishing emails and how to identify phishing URLs.
- According to learning science practice makes perfect and training games have to include practice questions. It is important that the feedback on the answers from the learner explain why something is right or wrong, instead of just showing correctness (Clark & Mayer, 2016). Every instruction video in PHREE is supported by four multiple-choice practice questions that relate directly to the topic of the video. After each question users are presented with a detailed explanation about why each multiple-choice answer is right or wrong (figure 6).
- Finally, users should have a form of control in game-based training. Proper training should allow its user to determine his or her pace. Control can be exercised in three ways: (1) by content sequencing (when learners can control the order of the training materials), (2) by pacing (when learners can control the time spent per lesson), and (3) by giving the option of learning support (when learners can ask for instructions) (Clark & Mayer, 2016). In PHREE users can control the pace of learning by skipping instructional videos. Besides, a cheat sheet, available at every practice question, gives users the possibility to ask for extra support.

Figure 6: Procedures of PHREE



The story line of PHREE (Fun). The main character of PHREE is Bob Visvanger. Bob works for web shop TREET and the director of TREET is Mrs. Angry. The story line starts as follows: Bob made a huge mistake last month; he fell for a phishing attack which cost TREET a lot of money. Because of this, Mrs. Angry is furious and devises an appropriate punishment. Therefore, Mrs. Angry commands Bob to give four lectures (in the form of instructional videos) about phishing and, by doing so, prepare the four most vulnerable departments of the company (sales, marketing, administration, and ICT) for phishing attacks in the future. Each lecture has its own topic. After each lecture four employees will ask Bob a question that relates to the topic of the lecture (figure 5). Bob has a cheat sheet, which he can consult at all times, to answer each question correctly. However, Mrs. Angry is really upset and secretly keeps an eye on Bob's lectures to see if he answers questions incorrectly (figure 5). Mrs. Angry does not allow Bob to take more than one minute to answer a question nor does she allow him to answer more than two questions per round incorrect. If Bob fails

to meet these demands he will be fired and has to start over. Bob's goal is to get through all four lectures and corresponding questions without being fired (figure 6).

3.2 Pilot Study: Test PHREE

A pilot test was performed to test the effect of PHREE on the ability of users to identify the legitimacy of emails. The design was based on earlier anti-phishing experiments (Alnajim & Munro, 2009a; Kumaraguru et al., 2010; Sheng et al., 2007).

Emails. The phishing emails used in this study were based on real phishing emails as stored in the phishing database of Fraudehelpdesk, or adapted from legitimate emails just as phishers would do. All phishing emails contained at least one cue that revealed its fraudulent nature. The cue could be found in (1) the "from" field, (2) the content of the email, or (3) in links to websites.

Test moments. To test the effect of anti-phishing training users were tested for their ability to identify emails in three rounds. There was (1) a pretest, (2) a direct posttest, and (3) a retention test performed one week after training (Kumaraguru et al., 2010). The pretest was used to determine a base line level ability of users to correctly identify the legitimacy of emails. The direct posttest was to measure any effect directly after training. The retention test was to examine if any effect of training remained up to one week. Based on the study design of Sheng et al. (2007) in each round all users were presented 10 screenshots of emails containing five legitimate emails and five simulated phishing emails. All emails were addressed to Bob Visvanger, and participants were asked to imagine that was their own email address (all participants performed the three tests at home behind their own computer).

Treatment groups. There were two treatment conditions: "control" and "experimental". Users in the experimental condition received three rounds of PHREE training (for approximately 20 minutes) in between the pretest and the direct posttest. Users in the control condition remained untrained (table 3).

Table 3: Methodology of Experiment

Design	Pretest	Training	Direct posttest	Retention test
Control	C1		C2	C3
Experimental	E1	X	E2	E3

In total 30 emails were divided into three groups: email set A, B, and C. One-third of the participants (12 people) saw set A in the pretest and set B or C in the posttest. Another one-third of the participants (12 people) saw set B in the pretest and set A or C in the posttest. The last one-third of the participants (12 people) saw set C in the pretest and group A or B in the posttest (Sheng et al., 2007) (table 4).

Table 4: Distribution of Emails

Pretest	Direct posttest	Retention test
A	B	C
A	C	B
B	A	C
B	C	A
C	A	B
C	B	A

Research Design. In sum, the experiment went as follows. First, all participants were told that they participated in a research that examined the ability of users to identify the fraudulent nature of emails. Second, all participants saw 10 screenshots of emails and each participant was asked to state for each of these 10 emails if it was legitimate or phishing and how confident they were with their decision-making (Sheng et al., 2007). By asking all participants exactly what they were tested on (to judge emails as phishing or legitimate) both treatment groups were equally aware about the purpose of the study. This prevented a Hawthorne-effect in which trained users would perform better than untrained users in identifying phishing emails solely because training would make users aware that they are tested on phishing detection abilities. Thirdly, each participant in the experimental group received 20 minutes of PHREE anti-phishing training using PowerPoint. Fourth, users had to judge the legitimacy of 10 more emails and state how confident they were with each decision in the direct posttest. User in the control condition performed the posttest directly after the pretest. Users in the experimental condition performed the posttest directly after the third round of anti-phishing training. Finally, one week later all users were asked for the last time to identify 10 emails as legitimate or phishing and how confident they were with their decision making. Besides, trained users were asked to give open feedback on the training.

Recruitment and demographics. 36 participants were recruited via social media. The only precondition of this study was that participants had to have a minimum age of 18. Excel was used to randomly assign all participants over the treatment conditions and various email sets.

Data collection and measurement. By means of a questionnaire information was gathered about the effect of anti-phishing training, confidence of participants in their decision-making, and users' opinion on PHREE.

User performance was measured by three ratios (Alnajim & Munro, 2009a): (1) Total Correct Rate (TCR), the total percentage of correctly identified emails, (2) Phishing Rate (PR), the percentage of rightly recognized phishing emails, and (3) Legitimate Rate (LR), the percentage of correctly identified legitimate emails.

1. $TCR = \text{total correctly identified emails} / \text{total number of emails} (10)$
2. $PR = \text{correctly identified phishing emails} / \text{number of phishing emails} (5)$

3. LR = correctly identified legitimate emails / number of legitimate emails (5).

Confidence of users was measured by means of a Likert-scale, from 1 not confident at all to 5 very confident, for each email. Finally, after the last test, trained users were asked to give feedback on the anti-phishing training.

Data analyses. First, a two-way mixed ANOVA was used to examine if there was a main difference in performance (TCR, PR, or LR, confidence of users) between the two treatment groups (control and experimental) over time (pretest, direct posttest, and retention test). If there was a main effect, a repeated measurement ANOVA and independent *t*-testes were performed to analyze the mean performance differences within one group (control or experimental) over time, or between groups (control vs. experimental) at each time point. SPSS Statistics was used to perform all the statistical analyses. Finally, open feedback on PHREE made it possible to collect detailed and valuable information about the perceived quality of PHREE

4. RESULTS

This chapter describes the demographics and susceptibility of participants in this study as well as their performance during multiple rounds of anti-phishing tests.

4.1 Demographics

In total 36 participants were recruited and all users performed a pretest, direct posttest and a retention test. Not all participants were able to perform the retention test exactly one week after the direct posttest. Nonetheless, all users were able to perform the retention test approximately (six to eight days) one week after the direct posttest. 64% of all participants were male and 36% female. The youngest participant was 21 and the oldest 54 with a mean of 25.4. Except for one, all participants had an age between 21 and 32. Finally, 19 participants had (or pursued) a bachelor's degree, 12 participants a master's degree, and five participants had a practical school diploma (appendix 4, table 5 - table 7).

As users saw email set A ($n = 12$), emails set B ($n = 12$), or email set C ($n = 12$) during the pretest, a one-way ANOVA was used to determine if the three email sets were equally difficult. The results (appendix 4, table 8) revealed no statistically significant differences between email set A, B, or C in TCR [$F(2, 33) = 1.27, p = .29$], LR [$F(2, 33) = 1.70, p = .20$], or PR [$F(2, 33) = .99, p = .38$] scores.

4.2 User Performance

A Spearman's correlation was run to examine the relationship between gender, age or educational level with performance measured in TCR during the pretest (appendix 4, table 9). When considering the small sample size and an alpha level of 0.1, men performed slightly better than women $r(34) = -.28, p = .10$ in identifying the legitimacy of emails. Besides, people in their late twenties performed better than people in their early twenties $r(34) = .33, p = .05$ in identifying the fraudulent nature of emails (after removing one outlier participant with the age of 54). Finally, results indicated that there was no correlation between TCR and education level $r(34) = .15, p = .38$.

Since the demographic nature of both experimental groups was very similar (appendix 4, table 5 - table 7) and results were not statically significant influenced by the difference between email sets it was possible to analyze the interaction effect of *group* (experimental or control) and *time* (pretest, direct posttest, and retention test) on the dependent variables TCR, PR, and LR. A two-way mixed ANOVA was used because this is the appropriate method to analyze an interaction effect, with one within-subject factor *time* and one between-subject factor *group* and a continuous dependent variable TCR, PR, and LR (Laerd Statistics, 2015). The purpose of this interaction effect analyses was to determine if there were differences between trained users and untrained in identifying emails over time.

To test if the collected data could actually be analyzed by a two-way mixed ANOVA five assumptions of this method were considered for each dependent variable. (1) There were no outliers as assessed by examination of studentized residuals for values greater than ± 3 . (2) Data was approximately normally distributed according to Normal Q-Q plots. There was homogeneity of (3) variances ($p > .05$) and (4) covariance's ($p > .001$) as measured by Levene's test of homogeneity of variance and Box's M test. Finally, (5) Mauchly's test of sphericity indicated that the assumption of sphericity was met for the two-way interaction term *time*group* for each dependent variable [TCR: $\chi^2(2) = 1.33, p = .51$; PR: $\chi^2(2) = 2.56, p = .28$; LR: $\chi^2(2) = .45, p = .81$]. Since all assumptions were met, the two-way ANOVA was run to analyze the effect of anti-phishing training on each dependent variable.

Total correctly identified emails. The TCR score represented the total amount of correctly identified emails as phishing or legitimate. Figure 7 gives a visual representation of the TCR development for both treatment groups. Users in the control group judged approximately 72% of all emails correctly at each test moment. For users in the experimental condition the mean percentage of correctly identified emails was 68% before training, then increased to 86% after training, and dropped again to 83% during the retention test.

The two-way mixed ANOVA results indicated a statistically significant main interaction between group (control or experimental) and *time* (pretest, direct posttest, or retention test) on TCR, $F(2, 68) = 5.16, p = .01, \text{partial } \eta^2 = .13$ (appendix 4, table 10). Therefore, the ability to identify emails (phishing + legitimate) changed significantly differently over time depending on whether users were in the experimental group or in the control group (figure 7).

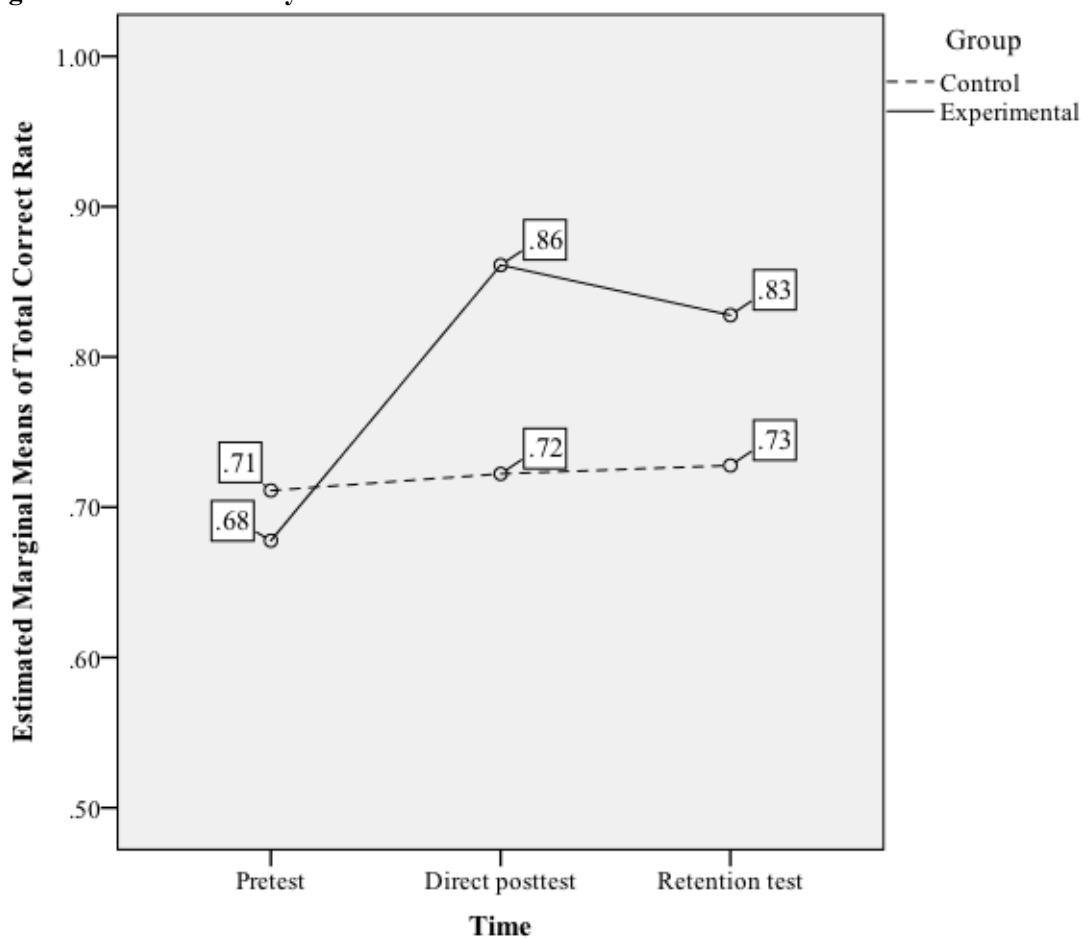
Interpreting main effects can be misleading; therefore, simple main effects need to be reported (Laerd Statistics, 2015). Analyzing simple main effects involves examining the statistical differences between the two treatment groups at each time moment and the differences within one treatment group over time.

Firstly, the differences between the two treatment groups were analyzed at each time moment. At the pretest there was not statistically significant difference in TCR scores between users in the experimental group and users in the control group $M = -0.03, SE = 0.05, t(34) = -.66, p = .52$.

However, at the direct posttest, there was a difference in TCR scores, with trained users scoring statistically significant higher than untrained users $M = 0.14$, $SE = 0.05$, $t(34) = 2.80$, $p = .01$. The difference in TCR remained and trained users still had higher TCR scores than untrained users after one week $M = 0.10$, $SE = 0.05$, $t(34) = 2.07$, $p = .05$.

Secondly, the difference in TCR scores within one group over three time points was examined. On the one hand there was no statistically significant effect of time on TCR for the control group $F(2, 34) = .10$, $p = .91$, partial $\eta^2 = .52$. On the other there was a statistically significant effect of time on TCR scores for the experimental group $F(2, 34) = 11.01$, $p < .01$, partial $\eta^2 = .39$. Trained users had statistically significant higher TCR scores at the direct posttest ($M = 1.83$, $SE = 0.04$ TCR, $p < .01$) and retention test ($M = 1.50$, $SE = 0.04$ TCR, $p = .01$) as compared to the pretest. Users in the experimental group performed significantly better after training than before. The difference in mean TCR score between the direct posttest and the retention test was not statistically significant ($M = -0.33$, $SE = 0.04$ TCR, $p = 1.00$). Users did not statistically significant loose their ability to identify emails correctly after one week compared to the direct posttest.

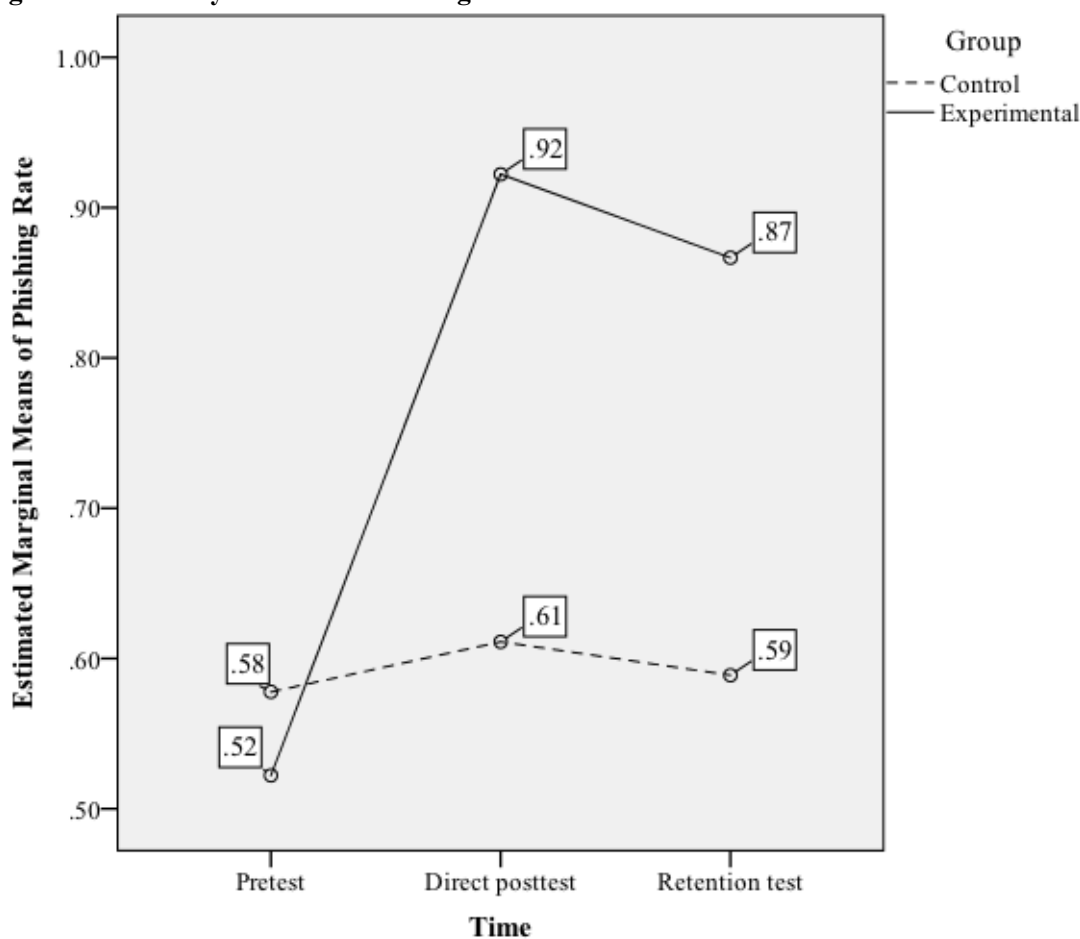
Figure 7: Total Correctly Identified Emails



Note: Total Correct Rate = total correctly identified emails / total number of emails (10)

Correctly identified phishing emails. Since the total amount of correctly identified emails increased for trained users, it was analyzed if this was due to increased PR or LR scores. The PR scores were of particular interest, because this rate represented the ability of users to identify phishing emails. Figure 8 gives a graphical representation of the development of PR scores for users in both experimental groups. The mean PR scores remained around 59% for users in the control group, but the mean PR scores changed over time for trained users and was 52% before training, 92% after training and 87% at the retention test.

Figure 8: Correctly Identified Phishing Emails



Note: Phishing Rate = correctly identified phishing emails / number of phishing emails (5)

The two-way mixed ANOVA results showed a statistically significant main interaction between *group* (control or experimental) and *time* (pretest, direct posttest, or retention test) on PR, $F(2, 68) = 11.56, p < .01, \text{partial } \eta^2 = .25$. Therefore, the ability to identify phishing emails changed significantly differently over the three time points depending on whether users were trained or remained untrained (appendix 4, table 11).

First, the differences between groups (control or experimental) were analyzed at each time moment. At the pretest there was no statistically significant difference in PR scores between users in

the experimental group and users in the control group $M = -0.05$, $SE = 0.08$, $t(34) = -.72$, $p = .48$. However, there was a statistically significant difference in PR scores between trained and untrained users at the direct posttest $M = 0.31$, $SE = 0.08$, $t(34) = 4.14$, $p < .01$ and the retention test $M = 0.28$, $SE = 0.06$, $t(34) = 4.37$, $p < .01$ with trained users scoring statistically significant higher than untrained users.

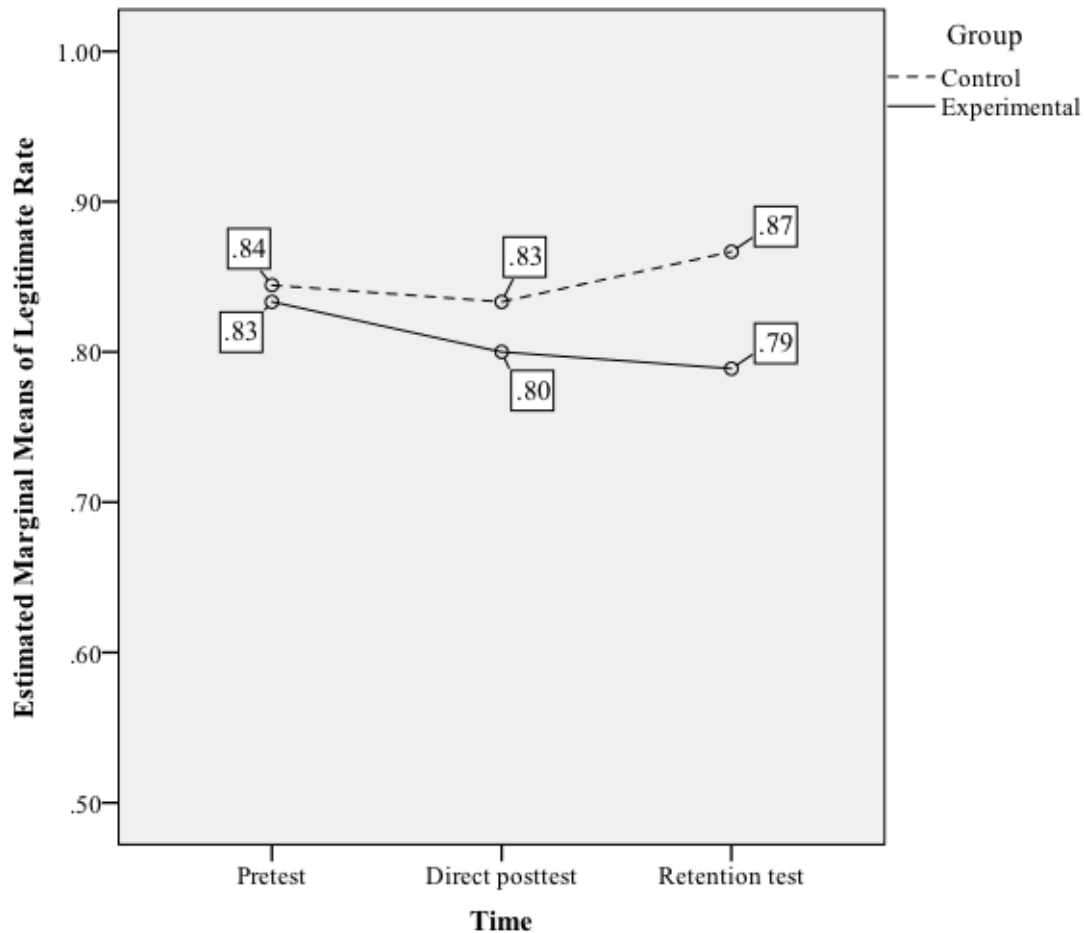
Secondly, the difference in PR scores within one group over three time points was examined. For users in the control group there was no statistically significant effect of time $F(2, 34) = .16$, $p = .85$, partial $\eta^2 = .01$. Conversely, there was a significant effect of time on PR scores for users in the experimental group $F(2, 34) = 26.54$, $p < .01$, partial $\eta^2 = .61$. Trained users had statistically significant greater ability to identify phishing emails in the direct posttest ($M = 0.40$, $SE = 0.06$ PR, $p < .01$) and retention test ($M = 0.34$, $SE = .06$ PR, $p < .01$) compared to the pretest. Consequently, users were significantly better able to identify phishing emails after training than before training. Finally, the difference in PR scores between the direct posttest and retention test decreased from $M = 0.92$ to $M = 0.87$, but this was insignificant ($M = -0.06$, $SE = 0.06$ PR, $p = 1.00$). Therefore, users did not statistically significant loose the ability to identify phishing emails after one week compared to the direct posttest.

Correctly identified legitimate emails. LR scores represented the ability of users to identify legitimate emails. Figure 9 shows that LR scores remained on a constant level for users in both experimental groups. Users in the control condition estimated around 85% of the legitimate emails correctly in each test round and users in the experimental condition identified approximately 80% correct at each test moment.

There was no main interaction effect between *group* (control or experimental) and *time* (pretest, direct posttest, or retention test) on LR $F(2, 68) = .40$, $p = .68$, partial $\eta^2 = .01$ (appendix 4, table 12). Users did not perform significantly different in LR scores over time depending in which treatment group they were. Therefore, the main effects for the within-subject factor *time* and the between-subject factor *group* were analyzed separately as suggested by Laerd Statistics (2015).

Firstly the effect of *group* was analyzed to examine if users in the experimental condition performed different from users in the control group over time regardless of a specific time point. Results showed there was no statistically significant difference in LR scores between the two treatment groups $F(1, 34) = .65$, $p = .43$, partial $\eta^2 = .02$.

Secondly the effect of *time* was analyzed to find out if performance in LR scores was different over time regardless of the treatment group. Hence, it was analyzed if pretest, direct posttest, and retention test LR scores differed from each other regardless of the treatment group. Results showed no statistically significant main effect of *time* $F(2, 68) = .17$, $p = .84$, partial $\eta^2 = .01$ (appendix 4, table 12).

Figure 9: Correctly Identified Legitimate Emails

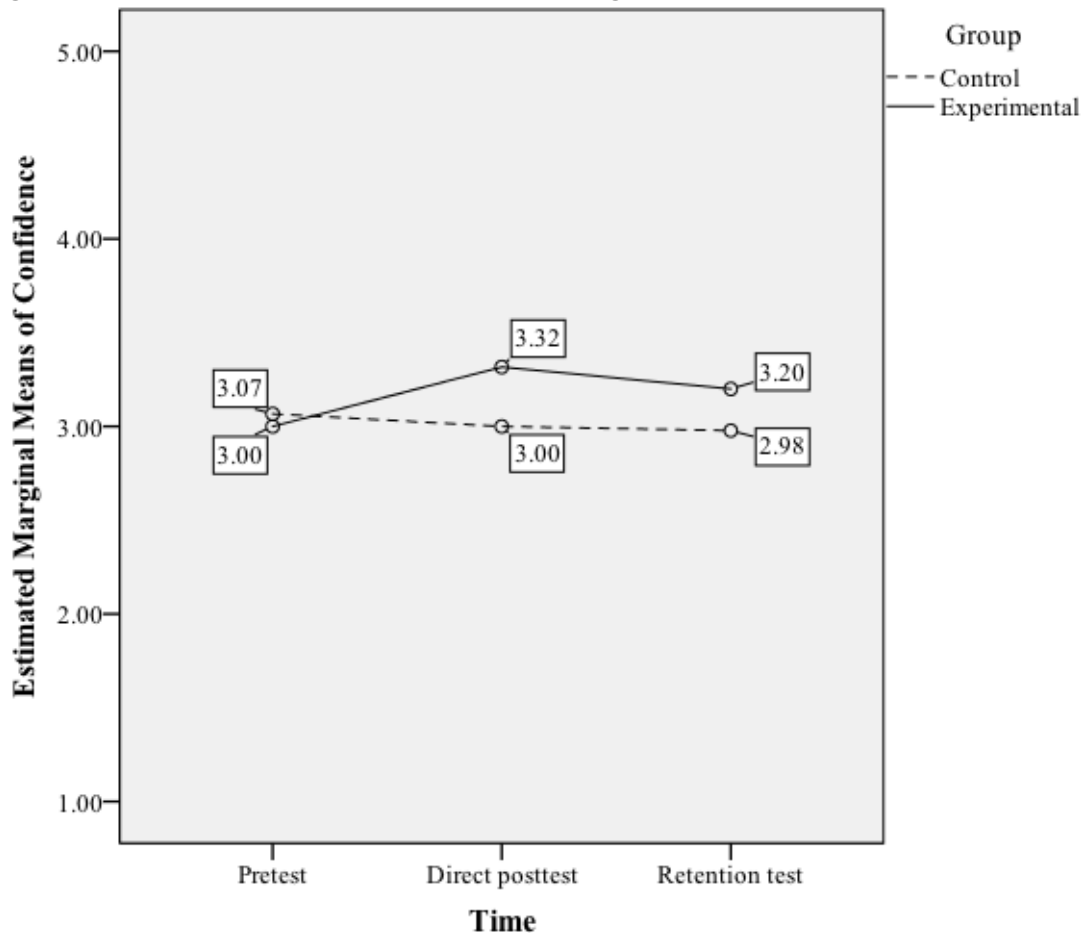
Note: Legitimate Rate = correctly identified legitimate emails / number of legitimate emails (5).

Confidence of users. For every email that participants judged, they also stated how confident they were in their decision-making from 1 not sure at all to 5 very sure. Users in the control group had a confidence score of approximately 3 at each test moment. Users in the experimental condition estimated their confidence with 3.07 before training, 3.32 directly after training, and 3.2 during the retention test (figure 10). Again a two-way mixed ANOVA was run to measure a main interaction effect. Results indicated that there was no interaction effect between *group* (control or experimental) and *time* (pretest, direct posttest, or retention test) on confidence of users $F(1.66, 65.69) = 3.02, p = .07, \text{partial } \eta^2 = .08$ (appendix 4, table 13). The confidence of users in their judgment over time was not significantly related to training. Therefore, the main effects for the within-subject factor *time* and the between-subject factor *group* were analyzed separately as suggested by Laerd Statistics (2015).

Firstly, there was no statistically significant effect of *group* on the confidence of users $F(1, 34) = 1.21, p = .28, \text{partial } \eta^2 = .03$. There was no statistically significant difference in confidence of users between the experimental group and users in the control group over time, regardless of a specific time point.

Secondly, there was no statistically significant effect of *time* $F(1.66, 65.69) = 1.19, p = .31$, partial $\eta^2 = .03$ (appendix 4, table 13). There was no statistically significant difference in confidence of users between the three time points when neglecting the treatment groups.

Figure 10: Confidence of Users in Decision-Making



4.3 User Feedback

After the training, users were asked to give their opinion on training PHREE. 17 users commented, from which 16 responses were positive. The most negative comment was *"nice, but I am not convinced I will never fall for a phishing attack in the future"*. The positive comments differed from short positive feedback as *"fun"*, *"interesting"*, and *"useful"* to more detailed feedback as *"I think this training can really help to solve a lot of problems at companies"*. One participant commented on the process of reading an email to judge its trustworthiness, *"Before training I paid attention to the type of company mentioned in the email. For example, ING seems like a trustworthy company, so I trusted that email to be legitimate. After the training I read the emails in a complete different way."* Another participant commented on the interactivity of the training: *"Very clear and interactive training. The idea to combine short lessons in a video with related questions afterwards forces you to immediately apply your newly gained knowledge."*

5. DISCUSSION AND CONCLUSION

In this thesis an anti-phishing training was developed and tested based on previous anti-phishing research. This chapter describes the key findings, limitations, future research, practical implications and conclusion.

5.1 Discussion

The developed anti-phishing training PHREE is highly effective and strongly increases the ability of users to identify phishing emails. This means that anti-phishing training with certain characteristics can effectively enhance the ability of users to identify phishing emails.

The main result confirmed previous studies on the effect of anti-phishing training (e.g. Canova et al., 2015; Kumaraguru, Rhee, Acquisti, et al., 2007) and also found that anti-phishing training can effectively enhance the ability of users to identify phishing emails.

An effective training is interactive (Sercombe & Papadaki, 2012), embedded (Kumaraguru, Rhee, Acquisti, et al., 2007), repetitive (Kumaraguru, Cranshaw, et al., 2009), with a game-based design (Sheng et al., 2007), in which text is kept simple and short and supported by graphics, and preferably has a cartoon format (Kumaraguru, Rhee, Acquisti, et al., 2007; Kumaraguru et al., 2008). Besides, the content of anti-phishing training should combine tips and tricks to recognize phishing emails (e.g. Kumaraguru, Cranshaw, et al., 2009) as well as phishing URLs (e.g. Canova et al., 2015) since phishing emails are often linked to phishing websites.

The results do not agree with other anti-phishing training studies that did not find a significant improvement directly after training or did not have statistically significant retention effect (Caputo et al., 2014; Kearney & Kruger, 2013). Since PHREE includes features of effective anti-phishing training, and excluded ineffective features, it may explain the positive results during the performed pilot test.

This study reveals a few more specific findings related to anti-phishing training PHREE. Firstly, trained users retain their ability to identify phishing emails for at least one week. Hence, users perform significantly better in identifying phishing emails one week after training than before training and users do not perform significantly worse at a retention test as compared to a direct posttest. A conclusion also found by some (e.g. Kumaraguru, Rhee, Sheng, et al., 2007), but not by others (e.g. Alnajim & Munro, 2009b). Therefore, the ability of users to retain anti-phishing knowledge depends on the characteristics of the training.

Secondly, training users to identify phishing emails does not negatively influence their ability to identify legitimate emails. However, the ability to identify legitimate emails reduced marginally after training. This is similar to Alnajim and Munro (2009a) their findings, who found that most training techniques do not change the ability to distinguish legitimate emails. However, other studies did find an improvement after training in the ability of users to identify legitimate emails (Kumaraguru et al., 2010). A possible explanation could be that PHREE teaches users to identify phishing emails, but not legitimate emails and, therefore, users recognize a phishing email when they

see one, but still are not sure how to distinguish a legitimate email. Another reason could be that trained users have a better understanding of the danger that phishing may pose to them, which may make users to decide to be on the safe side. Hence, after training users are more likely to classify a legitimate email as phishing because they fear the opposite of classifying a phishing email as legitimate.

Thirdly, training users by PHREE positively influences the ability of users to identify the total amount of emails (phishing + legitimate). This increased total ability to identify emails remains apparent for at least one week. The enhanced total performance reflects an enhanced ability of trained users to identify phishing emails and an unchanged ability of users to identify legitimate emails. An increase in phishing detection abilities with an unchanged ability to identify legitimate emails reflects some literature (Alnajim & Munro, 2009a), but not others (Kumaraguru et al., 2010).

Fourthly, users become slightly, although not significantly, more confident in their decision-making after training. This result supports earlier findings of minor improvement in confidence of users after training (Sheng et al., 2007). A possible explanation for the insignificant increase in confidence of users in their decision-making is that users are already moderately confident before training scoring on average a 3 out of 5 on a 1 (not at all confident) to 5 (very confident) scale. Another explanation could be that phishing emails in this experiment included only one indicator of phishing (sender, content, or link) per email. This may make it difficult for users to be a 100% sure about the legitimacy of an email, even after training. In real life phishing emails often contain multiple indicators of phishing.

Finally, no statistically significant effect of education level ($p > 0.1$) on the ability to identify emails correctly was measured. However, men performed marginally significant ($p < 0.1$) better than women in identifying the legitimacy of emails and people in their late twenties performed marginally significant ($p < 0.1$) better than people in their early twenties. The small difference in the ability to identify phishing emails between various demographic groups is a proper reflection of the literature. Some studies found that females indeed are more susceptible for phishing than males (Sheng, Holbrook, Kumaraguru, Cranor, & Downs, 2010) and users in the age group of 18 to 25 more vulnerable for phishing than users in other age groups (Sheng et al., 2010). However, others (e.g. Mohebzada et al., 2012) found no differences between demographic groups and phishing susceptibility.

Limitations. The experimental design and training have four limitations. Firstly, although the results in this thesis show that the training strongly improves the ability of users to identify phishing emails, it was a pilot test only. A small population of people in their twenties was recruited via social media. Besides, all users were aware that they participated in a research on the ability of users to distinguish phishing emails from legitimate emails. On the one hand this means that it is yet unknown how effective the training is in a (corporate) real-world setting. On the other hand, a pilot test is a first step in the process of developing a new training.

Secondly, the training was not embedded nor included repetition. To further improve the training it should be developed professionally, have an embedded design (Kumaraguru, Rhee, Acquisti, et al., 2007), and include repetition (Kumaraguru, Cranshaw, et al., 2009). This may increase the already positive effect of anti-phishing training on the ability of users to identify the legitimacy of emails.

Thirdly, the content of PHREE teaches users to identify current anti-phishing emails. However, phishing evolves and criminals will find new ways to gain sensitive information from users through phishing. This means that the content of PHREE needs to be flexible and must follow phishing trends in the future.

Finally, this study only considers one option in the form of training to solve the problem of phishing. However, lately other anti-phishing protection measures have established. For example banks do not use email as a communication tool anymore. Rather banks communicate directly to their customers with their own online platforms. However, it seems unlikely that all companies will use such an internal system to communicate with their customers and employees in the near future. Therefore, anti-phishing training remains necessary.

Future research. Several recommendations for follow-up studies derive from this thesis. Firstly, the results of this anti-phishing training pilot test are great. Therefore, it would be interesting to examine if training by PHREE could lead to less victims of phishing in a real-world (corporate) environment. If so, this training could be made publicly available, for example by publishing it on the website of Fraudehelpdesk, and actually help users to behave more secure online.

Secondly, although not statistically significant, a decrease of performance is visible between the direct posttest and retention tests. More research is needed to identify how long users will retain gained anti-phishing knowledge and actually keep on identifying phishing emails. It would also be interesting to know what effectively can limit the decrease of knowledge over time

Finally, multiple anti-phishing materials have been tested. Since PHREE has included most characteristics of effective training, it would be interesting to see if it is more effective than other developed training materials in teaching users to identify phishing emails.

Practical implications. Next to recommendations for further scientific research, some practical implications can also be derived from this study. Firstly, this thesis provides information on content and design characteristics that are central to the effectiveness of anti-phishing training. Therefore, if companies or individuals want to develop their own training, or have to choose between several options when buying a training, they now know more about what features the training should contain.

Secondly, the developed and tested anti-phishing training PHREE shows to be a very effective way to enhance the ability of users to identify phishing emails. Therefore, PHREE can effectively help individuals that want to increase their anti-phishing knowledge.

Finally, organizational leaders should understand that training can be an effective tool, according to published experiments, to prevent phishing damage. Hence, each company should carefully weigh the option of investing in a good training that may prevent damage or not to invest in training and taking the risk that sensitive (corporate) information may become public.

5.2 Conclusion

In conclusion, this study suggests that PHREE training strongly enhances the ability of users to identify phishing emails. PHREE is a, game-based training with a cartoon format in which users are presented simple and short instructional videos. The training is interactive because after each instructional video users have to answer four, topic related, questions. The content of PHREE combines tips and tricks to recognize phishing emails and phishing URLs, and offers a proper emergency solution for uncertain situations. To test PHREE, an experimental group (PHREE training) was compared to a control group (no training) in distinguishing phishing emails from legitimate emails. Users that received PHREE training improved their ability to identify emails (phishing + legitimate) from 68% correct before training to 86% correct after training. PHREE training especially enhanced the ability of users to recognize phishing emails from 52% correct before training to 92% correct after training. Trained users performed significantly better than untrained users who identified approximately 72% of all emails (phishing + legitimate) and 59% of the phishing emails correct at each test moment. The enhanced ability of trained users to identify phishing emails remained apparent for at least one week, without decreasing their ability to distinguish legitimate emails. Therefore, PHREE contains characteristics that make the training effective in teaching users to identify phishing emails. Overall the results strongly support the use PHREE as anti-phishing training. However, it must be noted that this experiment was a pilot test only. Further research is hence needed to determine the effect of PHREE on the ability of users to distinguish phishing emails from legitimate emails in a real-world (corporate) setting.

REFERENCES

- Aaron, G., & Manning, R. (2016). Phishing Activity Trends Report, 1st Quarter 216. Retrieved from <http://www.antiphishing.org/resources/apwg-reports/>
- Abbasi, A., Mariam Zahedi, F., & Chen, Y. (2016). *Phishing susceptibility: The good, the bad, and the ugly*. Paper presented at the 14th IEEE International Conference on Intelligence and Security Informatics, ISI 2015.
- Aburrous, M., Hossain, M. A., Dahal, K., & Thabtah, F. (2010). Experimental Case Studies for Investigating E-Banking Phishing Techniques and Attack Strategies. *Cognitive Computation*, 2(3), 242-253. doi:10.1007/s12559-010-9042-7
- Aggarwal, A., Kumar, V., & Sudarsan, S. D. (2014). *Identification and detection of phishing emails using natural language processing techniques*. Paper presented at the ACM International Conference Proceeding Series.
- Aggarwal, A., Rajadesingan, A., & Kumaraguru, P. (2012). *PhishAri: Automatic realtime phishing detection on twitter*. Paper presented at the eCrime Researchers Summit, eCrime.
- Ahmed, A. A., & Abdullah, N. A. (2016). *Real time detection of phishing websites*. Paper presented at the 7th IEEE Annual Information Technology, Electronics and Mobile Communication Conference, IEEE IEMCON 2016.
- Albladi, S., & Weir, G. R. S. (2016). *Vulnerability to social engineering in social networks: A proposed user-centric framework*. Paper presented at the 4th IEEE International Conference on Cybercrime and Computer Forensic, ICCCF 2016.
- Alnajim, A., & Munro, M. (2009a). *An anti-phishing approach that uses training intervention for phishing websites detection*. Paper presented at the 6th International Conference on Information Technology: New Generations, ITNG 2009, Las Vegas, NV.
- Alnajim, A., & Munro, M. (2009b). *An evaluation of users' anti-phishing knowledge retention*. Paper presented at the Proceedings - 2009 International Conference on Information Management and Engineering, ICIME 2009.
- Aloul, F. A. (2010). *Information security Awareness in UAE: A survey paper*. Paper presented at the 2010 International Conference for Internet Technology and Secured Transactions, ICITST 2010, London.
- Anderson, R., Barton, C., Böhme, R., Clayton, R., Van Eeten, M. J., Levi, M., . . . Savage, S. (2012). Measuring the cost of cybercrime. In R. Böhme (Ed.), *The Economics of Information Security and Privacy* (pp. 265-300): Springer Berlin Heidelberg.
- Arachchilage, N. A. G., & Cole, M. (2011). *Design a mobile game for home computer users to prevent from "phishing attacks"*. Paper presented at the International Conference on Information Society, i-Society 2011, London.
- Arachchilage, N. A. G., & Love, S. (2013). A game design framework for avoiding phishing attacks. *Computers in Human Behavior*, 29(3), 706-714.

- Arachchilage, N. A. G., Love, S., & Beznosov, K. (2016). Phishing threat avoidance behaviour: An empirical investigation. *Computers in Human Behavior, 60*, 185-197.
doi:10.1016/j.chb.2016.02.065
- Aston, M., McCombie, S., Reardon, B., & Watters, P. (2009). *A preliminary profiling of internet money mules: An australian perspective*. Paper presented at the Symposia and Workshops on Ubiquitous, Autonomic and Trusted Computing in Conjunction with the UIC'09 and ATC'09 Conferences, UIC-ATC 2009, Brisbane.
- Basnet, R., Mukkamala, S., & Sung, A. H. (2008) Detection of phishing attacks: A machine learning approach. *Vol. 226. Studies in Fuzziness and Soft Computing* (pp. 373-383).
- Bergholz, A., De Beer, J., Glahn, S., Moens, M. F., Paaß, G., & Strobel, S. (2010). New filtering approaches for phishing email. *Journal of Computer Security, 18*(1), 7-35. doi:10.3233/JCS-2010-0371
- Berthon, P., Pitt, L., & Watson, R. T. (1996). Marketing communication and the world wide web. *Business Horizons, 39*(5), 24-32.
- Besimi, A., Shehu, V., Abazi-Bexheti, L., & Dika, Z. (2009). *Managing security in a new learning management system (LMS)*. Paper presented at the ITI 2009 31st International Conference on Information Technology Interfaces, ITI 2009, Cavtat, Dubrovnik.
- Bowen, B. M., Devarajan, R., & Stolfo, S. (2011). *Measuring the human factor of cyber security*. Paper presented at the 11th IEEE International Conference on Technologies for Homeland Security, HST 2011, Waltham, MA.
- Canova, G., Volkamer, M., Bergmann, C., & Reinheimer, B. (2015). NoPhish app evaluation: lab and retention study.
- Caputo, D. D., Pfleeger, S. L., Freeman, J. D., & Johnson, M. E. (2014). Going spear phishing: Exploring embedded training and awareness. *IEEE Security and Privacy, 12*(1), 28-38.
doi:10.1109/MSP.2013.106
- Claffey Jr, G. F., & Regan, H. J. (2011). *InnovatEDU a collaboration to reduce higher ed security risk*. Paper presented at the 39th Annual ACM SIGUCCS Conference, SIGUCCS'11, San Diego, CA.
- Clark, R. C., & Mayer, R. E. (2016). *E-learning and the science of instruction: Proven guidelines for consumers and designers of multimedia learning*: John Wiley & Sons.
- Davinson, N., & Sillence, E. (2010). It won't happen to me: Promoting secure behaviour among internet users. *Computers in Human Behavior, 26*(6), 1739-1747.
doi:10.1016/j.chb.2010.06.023
- Dhamija, R., Tygar, J. D., & Hearst, M. (2006). *Why phishing works*. Paper presented at the Conference on Human Factors in Computing Systems - Proceedings.
- Dodge Jr, R. C., Carver, C., & Ferguson, A. J. (2007). Phishing for user security awareness. *Computers and Security, 26*(1), 73-80. doi:10.1016/j.cose.2006.10.009

- Dodge, R. C., Coronges, K., & Rovira, E. (2012) Empirical benefits of training to phishing susceptibility. *Vol. 376 AICT. 27th IFIP TC 11 Information Security and Privacy Conference, SEC 2012* (pp. 457-464). Heraklion, Crete.
- Dodge, R. C., & Ferguson, A. J. (2006) Using phishing for user email security awareness. *Vol. 201. IFIP International Federation for Information Processing* (pp. 454-459).
- Dong, X., Clark, J. A., & Jacob, J. (2008). *Modelling user-phishing interaction*. Paper presented at the 2008 Conference on Human System Interaction, HSI 2008.
- Downs, J. S., Holbrook, M. B., & Cranor, L. F. (2006). *Decision strategies and susceptibility to phishing*. Paper presented at the ACM International Conference Proceeding Series.
- Emigh, A. (2005). *Online identity theft: phishing technology, chokepoints and countermeasures*. Washington, DC: Identity Theft Technology Council.
- Falk, J. D., & Kucherawy, M. S. (2010). Battling spam: The evolution of mail feedback loops. *IEEE Internet Computing, 14*(6), 68-71. doi:10.1109/MIC.2010.133
- Flores, W. R., Holm, H., Nohlberg, M., & Ekstedt, M. (2015). Investigating personal determinants of phishing and the effect of national culture. *Information and Computer Security, 23*(2), 178-199. doi:10.1108/ICS-05-2014-0029
- Forte, D. (2009). Phishing in depth. *Network Security, 2009*(5), 19-20. doi:10.1016/S1353-4858(09)70055-8
- Frauenstein, E. D., & Von Solms, R. (2014). *Combatting phishing: A holistic human approach*. Paper presented at the 2014 Annual Conference on Information Security for South Africa, ISSA 2014.
- Gupta, S., & Kumaraguru, P. (2014). *Emerging phishing trends and effectiveness of the anti-phishing landing page*. Paper presented at the 2014 APWG Symposium on Electronic Crime Research, eCrime 2014.
- Halaseh, R. A., & Alqatawna, J. (2016). *Analyzing cybercrimes strategies: The case of phishing attack*. Paper presented at the Proceedings - 2016 Cybersecurity and Cyberforensics Conference, CCC 2016.
- Hale, M. L., Gamble, R. F., & Gamble, P. (2015). *CyberPhishing: A game-based platform for phishing awareness testing*. Paper presented at the 48th Annual Hawaii International Conference on System Sciences, HICSS 2015.
- He, M., Horng, S. J., Fan, P., Khan, M. K., Run, R. S., Lai, J. L., . . . Sutanto, A. (2011). An efficient phishing webpage detector. *Expert Systems with Applications, 38*(10), 12018-12027. doi:10.1016/j.eswa.2011.01.046
- Herzberg, A., & Jbara, A. (2008). Security and identification indicators for browsers against spoofing and phishing attacks. *Acm Transactions on Internet Technology, 8*(4). doi:10.1145/1391949.1391950
- Hong, J. (2012). The state of phishing attacks. *Communications of the ACM, 55*(1), 74.

- Internet Users. (2016). *Internet Live Stats*. Retrieved from <http://www.internetlivestats.com/internet-users/>
- Jagatic, T. N., Johnson, N. A., Jakobsson, M., & Menczer, F. (2007). Social phishing. *Communications of the ACM*, 50(10), 94-100. doi:10.1145/1290958.1290968
- Jansson, K., & Von Solms, R. (2011). *Simulating malicious emails to educate end users on-demand*. Paper presented at the IEEE Symposium on Web Society.
- Jansson, K., & Von Solms, R. (2013). Phishing for phishing awareness. *Behaviour and Information Technology*, 32(6), 584-593. doi:10.1080/0144929X.2011.632650
- Jeeva, S. C., & Rajsingh, E. B. (2016). Intelligent phishing url detection using association rule mining. *Human-Centric Computing and Information Sciences*, 6(1). doi:10.1186/s13673-016-0064-3
- Kearney, W. D., & Kruger, H. A. (2013) Phishing and organisational learning. *Vol. 405. IFIP Advances in Information and Communication Technology* (pp. 379-390).
- Kearney, W. D., & Kruger, H. A. (2014). *Considering the influence of human trust in practical social engineering exercises*. Paper presented at the Information Security for South Africa (ISSA), 2014.
- Kritzinger, E. (2016). Short-term initiatives for enhancing cyber-safety within South African schools. *South African Computer Journal*, 28(1), 1-17. doi:10.18489/sacj.v28i1.369
- Kumaraguru, P., Cranor, L. F., & Mather, L. (2009). *AntiPhishing landing page: Turning a 404 into a teachable moment for end users*. Paper presented at the 6th Conference on Email and Anti-Spam, CEAS 2009, Mountain View, CA.
- Kumaraguru, P., Cranshaw, J., Acquisti, A., Cranor, L., Hong, J., Blair, M. A., & Pham, T. (2009). *School of phish: A real-world evaluation of anti-phishing training*. Paper presented at the 5th Symposium On Usable Privacy and Security, SOUPS 2009, Mountain View, CA.
- Kumaraguru, P., Rhee, Y., Acquisti, A., Cranor, L. F., Hong, J., & Nunge, E. (2007). *Protecting people from phishing: The design and evaluation of an embedded training email system*. Paper presented at the 25th SIGCHI Conference on Human Factors in Computing Systems 2007, CHI 2007, San Jose, CA.
- Kumaraguru, P., Rhee, Y., Sheng, S., Hasan, S., Acquisti, A., Cranor, L. F., & Hong, J. (2007). *Getting users to pay attention to anti-phishing education: Evaluation of retention and transfer*. Paper presented at the Anti-Phishing Working Groups 2nd Annual eCrime Researchers Summit, eCrime '07, Pittsburgh, PA.
- Kumaraguru, P., Sheng, S., Acquisti, A., Cranor, L. F., & Hong, J. (2008). *Lessons from a real world evaluation of anti-phishing training*. Paper presented at the eCrime Researchers Summit, eCrime 2008, Atlanta, GA.
- 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). doi:10.1145/1754393.1754396

- Lastdrager, E. E. (2014). Achieving a consensual definition of phishing based on a systematic review of the literature. *Crime Science*, 3(1), 1-10.
- Lastdrager, E.E.H., Carvajal Gallardo, I., Hartel, P.H., & Junger, M. (2017) *How Effective is Anti-Phishing Training for Children?* Paper presented at SOUPS 2017: 13th Symposium on Usable Privacy and Security, Santa Clara, California, USA. USENIX Association.
- Laerd Statistics (2015). Two-way mixed ANOVA using SPSS Statistics. Statistical tutorials and software guides. Retrieved from <https://statistics.laerd.com/>
- Lim, I. K., Park, Y. G., & Lee, J. K. (2016). Design of Security Training System for Individual Users. *Wireless Personal Communications*, 90(3), 1105-1120. doi:10.1007/s11277-016-3380-z
- Maignan, I., & Lukas, B. A. (1997). The Nature and Social Uses of the Internet: A Qualitative Investigation. *Journal of Consumer Affairs*, 31(2), 346-371.
- Marett, K., & Wright, R. (2009). *The effectiveness of deceptive tactics in phishing*. Paper presented at the 15th Americas Conference on Information Systems 2009, AMCIS 2009.
- Mayhorn, C. B., & Nyeste, P. G. (2012). Training users to counteract phishing. *Work-a Journal of Prevention Assessment & Rehabilitation*, 41(SUPPL.1), 3549-3552. doi:10.3233/WOR-2012-1054-3549
- Milne, G. R., & Gordon, M. E. (1993). Direct mail privacy-efficiency trade-offs within an implied social contract framework. *Journal of Public Policy & Marketing*, 206-215.
- Mohebzada, J. G., Zarka, A. E., Bhojani, A. H., & Darwish, A. (2012). *Phishing in a university community: Two large scale phishing experiments*. Paper presented at the 2012 International Conference on Innovations in Information Technology, IIT 2012.
- Norris, D., Joshi, A., & Finin, T. (2015). *Cybersecurity challenges to American state and local governments*. Paper presented at the Proceedings of the European Conference on e-Government, ECEG.
- Parsons, K., McCormac, A., Pattinson, M., Butavicius, M., & Jerram, C. (2015). The design of phishing studies: Challenges for researchers. *Computers and Security*, 52, 194-206. doi:10.1016/j.cose.2015.02.008
- Paul, P. (1996). Marketing on the Internet. *Journal of Consumer Marketing*, 13(4), 27-37.
- Robila, S. A., & Ragucci, J. W. (2006). *Don't be a phish: Steps in user education*. Paper presented at the Working Group Reports on ITiCSE on Innovation and Technology in Computer Science Education 2006.
- Sercombe, A. A., & Papadaki, M. (2012). *Education in the 'virtual' community: Can beating Malware Man teach users about social networking security?* Paper presented at the Proceedings of the 6th International Symposium on Human Aspects of Information Security and Assurance, HAISA 2012.
- Sheehan, K. B., & Hoy, M. G. (2000). Dimensions of privacy concern among online consumers. *Journal of Public Policy and Marketing*, 19(1), 62-73.

- Sheng, S., Holbrook, M., Kumaraguru, P., Cranor, L. F., & Downs, J. (2010). *Who falls for phish? A demographic analysis of phishing susceptibility and effectiveness of interventions*. Paper presented at the Conference on Human Factors in Computing Systems - Proceedings.
- Sheng, S., Magnien, B., Kumaraguru, P., Acquisti, A., Cranor, L. F., Hong, J., & Nunge, E. (2007). *Anti-Phishing Phil: The design and evaluation of a game that teaches people not to fall for phish*. Paper presented at the SOUPS 2007: 3rd Symposium On Usable Privacy and Security, Pittsburgh, PA.
- Smadi, S., Aslam, N., Zhang, L., Alasem, R., & Hossain, M. A. (2015). *Detection of phishing emails using data mining algorithms*. Paper presented at the 9th International Conference on Software, Knowledge, Information Management and Applications, SKIMA 2015.
- Smith, A., Papadaki, M., & Furnell, S. M. (2009). *Improving awareness of social engineering attacks*. Paper presented at the IFIP World Conference on Information Security Education.
- Song, Y., Yang, C., & Gu, G. (2010). *Who is peeping at your passwords at starbucks? - To catch an evil twin access point*. Paper presented at the 2010 IEEE/IFIP International Conference on Dependable Systems and Networks, DSN 2010, Chicago, IL.
- Steyn, T., Kruger, H. A., & Drevin, L. (2007). *Identity theft—Empirical evidence from a Phishing exercise*. Paper presented at the IFIP International Information Security Conference.
- Stikic, M., Berka, C., & Korszen, S. (2015) Neuroenhancement in tasks, roles, and occupations. *Vol. 7. Monographs in Leadership and Management* (pp. 169-186): Emerald Group Publishing Ltd.
- Stockhardt, S., Reinheimer, B., Volkamer, M., Mayer, P., Kunz, A., Rack, P., & Lehmann, D. (2016) Teaching phishing-security: Which way is best? : *Vol. 471. 31st IFIP TC 11 International Conference on Systems Security and Privacy Protection, SEC 2016* (pp. 135-149): Springer New York LLC.
- Sun, J. C. Y., & Lee, K. H. (2016). Which teaching strategy is better for enhancing anti-phishing learning motivation and achievement? The concept maps on tablet PCs or worksheets? *Educational Technology and Society*, 19(4), 87-99.
- Tyler, J. (2016). Don't be your own worst enemy: protecting your organisation from inside threats. *Computer Fraud and Security*, 2016(8), 19-20. doi:10.1016/S1361-3723(16)30063-X
- Welk, A. K., Hong, K. W., Zielinska, O. A., Tembe, R., Murphy-Hill, E., & Mayhorn, C. B. (2015). Will the "Phisher-men" reel you in? assessing individual differences in a phishing detection task. *International Journal of Cyber Behavior, Psychology and Learning*, 5(4), 1-17. doi:10.4018/IJCBPL.2015100101
- Wolfswinkel, J. F., Furtmueller, E., & Wilderom, C. P. (2013). Using grounded theory as a method for rigorously reviewing literature. *European Journal of Information Systems*, 22(1), 45-55.

- Xiang, G., Hong, J., Rose, C. P., & Cranor, L. (2011). CANTINA+: A feature-rich machine learning framework for detecting phishing web sites. *Acm Transactions on Information and System Security*, 14(2). doi:10.1145/2019599.2019606
- Yang, C.-C., Tseng, S.-S., Lee, T.-J., Weng, J.-F., & Chen, K. (2012). *Building an anti-phishing game to enhance network security literacy learning*. Paper presented at the 2012 IEEE 12th International Conference on Advanced Learning Technologies.
- Zhang, Y., Egelman, S., Cranor, L., & Hong, J. (2007). *Phinding phish: Evaluating anti-phishing tools*. Paper presented at the 14th Annual Network and Distributed System Security Symposium (NDSS 2007), San Diego, CA.
- Zhang-Kennedy, L., Fares, E., Chiasson, S., & Biddle, R. (2016). *Geo-Phisher: The design and evaluation of information visualizations about internet phishing trends*. Paper presented at the eCrime Researchers Summit, eCrime.
- Zielinska, O. A., Tembe, R., Hong, K. W., Ge, X., Murphy-Hill, E., & Mayhorn, C. B. (2014). *One phish, two phish, how to avoid the internet phish: Analysis of training strategies to detect phishing emails*. Paper presented at the 58th International Annual Meeting of the Human Factors and Ergonomics Society, HFES 2014.

APPENDICES

Appendix 1: Final Selection Literature Review in Chronological Order

Literature that examines the effect of anti-phishing training.		
Author	Experimental design	Results
(Dodge Jr et al., 2007)	Unknowing students from the United States Military Academy were sent phishing to their student emails. $N = 512$ (test 1); $N = 4,118$ (test 2), $N = 4,136$ (test 3).	<ul style="list-style-type: none"> • Approximately 51% of the freshman, 31% of the sophomores, 23% of the juniors, and 18% of the seniors fell for phishing. • Two classes participated in all three experiments. The percentage that fell for phishing was 91% and 84% at test one, 39% and 44% at test two, and 30% and 24% at test three.
(Kumaraguru, Rhee, Acquisti, et al., 2007)	Participants recruited around Carnegie Mellon University interacted with an email inbox containing 19 emails. Email 3, 14, 16, and 17 were phishing emails and emails 5 and 11 were training emails. Three treatment groups: text and graphics ($n = 10$), comic strip ($n = 10$), security notices ($n = 10$).	<ul style="list-style-type: none"> • In the security notice group, 90% fell for phishing before training and 63%, on average over the three phishing emails, after training. From the 90% that fell for the first phishing email, 89% fell for the final phishing email. • In the text and graphics group, 80% fell for phishing before training and 30%, on average over the three phishing emails, after training. From the 80% that fell for the first phishing email, 63% fell for the final phishing email. • In the comic strip group, 100% fell for phishing before training and 23%, on average over three phishing emails, after training. From the 100% that fell for the first phishing email, 30% fell for the final phishing email. • The comic strip group performed significantly better than the other treatment groups.
(Kumaraguru, Rhee, Sheng, et al., 2007)	Participants recruited around Carnegie Mellon University interacted with an email inbox containing 33 emails + 16 emails at a retention test after one week. Three treatment groups: embedded (PhishGuru, $n = 14$), non-embedded (Amazon, $n = 14$), Control (email from friend, $n = 14$).	<ul style="list-style-type: none"> • Users in the control group identified 7% of all emails (phishing or legitimate) correctly in the pretest, 11% in the direct posttest, and 7% during the retention test. • Users in the non-embedded condition identified 4% of all emails correctly in the pretest, 14% in the direct posttest, and 7% during the retention test. • Users in the embedded condition identified 18% of all emails correctly in the pretest, 68% in the direct posttest, and 64% during the retention test. • The embedded group performed significantly better than the other treatment groups.

Literature that examines the effect of anti-phishing training.		
Author	Experimental design	Results
(Sheng et al., 2007)	Participants recruited around university and via craigslist had to identify 10 websites before and 10 websites after training. Three treatment groups: existing training (eBay, $n = 14$), tutorial (Phishing Phil on paper, $n = 14$), and game (Phishing Phil, $n = 14$).	<ul style="list-style-type: none"> • Users in the existing training material group identified 66% of all websites (phishing or legitimate) correctly in the pretest and 74% during the posttest. • Users in the tutorial condition identified 65% of all websites correctly in the pretest and 80% during the posttest. • Users in the game condition identified 69% of all websites correctly in the pretest and 87% in the posttest. • The game condition performed significantly better than the existing training material group in the posttest. The difference between the tutorial condition and the existing training material condition was not significant (possibly because of the small N).
(Kumaraguru et al., 2008)	Unknowing employees from a large Portuguese company were sent four simulated phishing emails. Three treatment groups: control (not trained $n = 111$), Generic (PhishGuru, $n = 110$), Spear (PhishGuru, $n = 100$).	<ul style="list-style-type: none"> • 42% of the users in the generic training condition clicked and gave information in response to the first phishing email. From these users 19% gave information the posttest one day later, and 12% during the retention test after one week. • 39% of the users in the spear training condition clicked and gave information in response to the first phishing email. From these users 18% gave information the posttest one day later, and 15% during the retention test after one week. • There was no statistically significant difference between the generic and the spear training condition.
(Alnajim & Munro, 2009a)	Participants without knowledge about phishing had to interact with an email inbox and saw seven websites before training and six after training. Three treatment groups: control (no training, $n = 12$) old approach (tips via email, $n = 12$), new approach (APTIPWD, $n = 12$).	<ul style="list-style-type: none"> • Users in the new approach condition identified 52% of all emails (phishing or legitimate) correctly in the pretest and 77% during the posttest. • Users in the old approach condition identified 50% of all emails correctly in the pretest and 52% during the posttest. • Users in the control condition identified 52% of all emails correctly in the pretest and 52% in the posttest. • The new approach condition performed significantly better than the old approach condition and the control group in the posttest.

Literature that examines the effect of anti-phishing training.		
Author	Experimental design	Results
(Kumaraguru, Cranshaw, et al., 2009)	Unknowing email users from the Carnegie Mellon University were sent three legitimate and seven phishing emails over 28 days. Three treatment groups: control (no training, $n = 172$), single-training (PhishGuru at day 0, $n = 172$), and multiple-training (PhishGuru at day 0 and 14, $n = 171$).	<ul style="list-style-type: none"> • 54.4% of the users in the control group that clicked on the first phishing email ($n = 90$) also clicked on the phishing email on day 28. • 27% of the users in the single-training condition that clicked on the first phishing email ($n = 89$) also clicked on the phishing email on day 28. This percentage was 42.9% on day 16. • 32.5% of the users in the multiple-training condition group that clicked on the first phishing email ($n = 90$) also clicked on the phishing email on day 28. This percentage was 26.5% on day 16. • The multiple-training condition performed significantly better than the single-training condition on day 16 and day 21. There was no significant difference between the two groups on day 28. • Trained participants performed significantly better than untrained participants.
(Smith et al., 2009)	46 people (students staff) tested an anti-phishing websites (Social-Ed) that contained information and quizzes on phishing.	<ul style="list-style-type: none"> • 69% of the users that read the training material on the website, passed a phish related quiz. • 44% of users that did not read the training material on the website, passed a phish related quiz.
(Aburrouss et al., 2010)	Employees at a bank (UK) had to identify 50 websites as legitimate or fraudulent. Two conditions: trained (did not receive training, but were in a prior phishing test, $n = 50$) and untrained ($n = 50$).	<ul style="list-style-type: none"> • Employees that were confronted with phishing before were better in distinguishing phishing websites from legitimate websites (72% correct) than people without experience (72% wrong).

Literature that examines the effect of anti-phishing training.		
Author	Experimental design	Results
(Davinson & Sillence, 2010)	Students were asked how secure they behaved online last week, would do in the next week, and actually did one week later. Two conditions: trained (Anti-Phishing Phil, $n = 32$), not trained ($n = 32$).	<ul style="list-style-type: none"> • The study did not find any difference between trained students and untrained users. Everyone acted more secure. • This may have two reasons. (1) Behavior was not monitored. (2) All students were warned to behave secure.
(Kumaraguru et al., 2010)	Participants recruited through an online mailing had to identify 18 websites, six before training, six directly after, and six one week later. Control (no training, $n = 2,496$), Game (Anti-Phishing Phil, $n = 2,021$ and $n = 674$ at the retention test.).	<ul style="list-style-type: none"> • Users who scored poorly in the pretest (maximum four websites were estimated correctly) improved their ability to identify phishing websites significantly. • In the pretest 57% of the phishing websites were identified as legitimate (false negatives), in the direct posttest this percentage dropped to 22%, and stayed 22% during the delayed posttest. • The study does not mention if the control condition (saw 12 websites) improved in the posttest as compared to the pretest.
(Bowen et al., 2011)	Unknowing students and staff received simulated phishing emails. People that fell for the attack received a warning message and were sent a new phishing email one week later. There were two experiments ($N = 500$, and $N = 2,000$)	<ul style="list-style-type: none"> • At the first experiment 313 users fell for the first phishing email, from these 313 users, 21 users fell for the second phishing email, from these 21 users only one user fell for the third phishing email, and no one fell for the fourth phishing email. • At the second experiment 384 users fell for the first phishing email, from these 384 users, 29 users fell for the second phishing email, from these 29 users only four users fell for the third phishing email, and no one fell for the fourth phishing email.

Literature that examines the effect of anti-phishing training.		
Author	Experimental design	Results
(Dodge et al., 2012)	Unknowing students were sent three simulated phishing emails. The first test determined a base level. A retention tests was performed after 10 days, and after six weeks. All training was embedded. Three treatment groups: no notification (error, $n = 287$), notification (feedback, $n = 298$), training (awareness training, $n = 307$).	<ul style="list-style-type: none"> • 56% of the users that were in the no notification group fell for the first phishing attempt, approximately 9% for the second attempt, and 47.5% for the last attempt. • 46% of the users that were in the feedback condition fell for the first phishing attempt, approximately 10% for the second attempt, and 32.08% for the last attempt. • 42% of the users that were in the feedback condition fell for the first phishing attempt, approximately 9% for the second attempt, and 24.5% for the last attempt. • Training had the most effect on phish avoidance behavior, than feedback, and an error message had the least effect.
(Yang et al., 2012)	62 Taiwanese students had to identify 20 websites as phishing or legitimate before and after playing the Anti-Phishing Education Game. A control group did not receive training ($N = \text{unknown}$).	<ul style="list-style-type: none"> • There is a significant improvement after playing the Anti-Phishing Education Game (paired t-test < 0.0001). • The control group also made significant improvements in the posttest as compared to the pretest (paired t-test $p < 0.01$).
(Mayhorn & Nyeste, 2012)	Participants recruited from the North Carolina State University psychology pool had to interact with an email inbox with 30 emails, and 40 in a retention test after one week. Three treatment groups: game and embedded (Anti-Phishing Phil + cartoon, $n = 28$), embedded (cartoon, $n = 28$), and control (no training, $n = 28$).	<ul style="list-style-type: none"> • In the direct posttest the game condition and the embedded conditions performed significantly better than the control group in identifying emails (factors repeated measures ANOVA, $p < 0.01$). • There was no significant difference in performance between the game condition and the embedded condition (Turkey HSD post hoc test). • There was no significant retention effect after one week.

Literature that examines the effect of anti-phishing training.		
Author	Experimental design	Results
(Jansson & Von Solms, 2013; Mayhorn & Nyeste, 2012)	25,579 unknowing students at Nelson Mandela Metropolitan University (South Africa) received two simulated phishing emails from which the first contained a link to an embedded warning message.	<ul style="list-style-type: none"> • 14.06% of users of the active email users ($N = 9,273$) reacted to the first phishing email. • 8.06% of the active email users ($N = 8,231$) reacted to the second phishing email. • Based on the difference in active users there were 42.63% less reactions in week two. • In total 976 users fell for phishing in week one, but not in week two, while being active email users in both weeks. Therefore, 11.85% of the total population learned from training.
(Caputo et al., 2014)	1,500 participants were randomly selected out of 6,000 employees from medium-sized Washington, DC-based firm and sent three simulated phishing emails. A baseline measurement and two retention tests. Two treatment groups: trained (two-column text training, $n = 1,078$) and untrained (received a warning message, $n = 281$).	<ul style="list-style-type: none"> • On average 62% of the untrained users fell for the first phishing attempt, and 36% for the second attempt. • On average 60% of the trained users fell for the first phishing before training and 35% after training. • 11% of the users clicked on all links regardless of their training condition. • 22% of the users did not click on links regardless of their training condition. • Participants did not read the training page and so did not actually receive training.
(Gupta & Kumaraguru, 2014)	3,359 unique IP-addresses were tracked. The data was obtained through a data set. The study analyzed, among many other things, if users learned from the anti-phishing landing page.	<ul style="list-style-type: none"> • The dataset revealed that users clicked 46% less often on blacklisted sites in April 2011 as compared to January 2011, indicating that the training worked.
(Sercombe & Papadaki, 2012)	104 students and employees from the Plymouth University had to answer a phish related survey. Two treatment groups: control (no training); Malware Man (game training, N per group unknown).	<ul style="list-style-type: none"> • Trained users answered 77% of the phishing survey questions correctly. • Untrained users answered 55% of the phishing survey questions correctly.

Literature that examines the effect of anti-phishing training.		
Author	Experimental design	Results
(Zielinska et al., 2014)	96 users (recruited via Mechanical Turk) received anti-phishing training. In addition one group saw a cooking video (control group, $n = 32$), a second group saw a video on the consequences of phishing (vignettes of loss, $n = 32$), and a third group saw news articles that aimed to increase the general fear level (trust section, $n = 32$).	<ul style="list-style-type: none"> • An analysis of variance showed that there was no significant difference between the treatment groups before and after training. • Anti-phishing training enhanced the ability of users to identify emails correctly ($t(95) = 3.01, p < .01$).
(Canova et al., 2015)	19 participants recruited via social media and flyers had to identify 16 websites before playing NoPhish, 16 after playing NoPhish and 16 at a retention test after five months.	<ul style="list-style-type: none"> • Before training users identified 57% of the URLs correctly. • At the direct posttest users identified 90% of the URLs correctly. • At the retention test users identified 81% of the URLs correctly. • Users significantly performed better at the posttest and the retention test as compared to the baseline measurement. • Users significantly lost knowledge during the retention test compared to the direct posttest.
(Arachchilage et al., 2016)	20 participants recruited at Brunel University had to identify 20 websites before playing the mobile version of Anti-Phishing Phil, and 20 websites afterwards.	<ul style="list-style-type: none"> • Before playing the game participants identified 56% of the websites correctly. • After playing the game participants identified 80% of the websites correctly.

Literature that examines the effect of anti-phishing training.		
Author	Experimental design	Results
(Stockhardt et al., 2016)	81 participants recruited at a school received similar training content delivered via different designs and had to identifying 16 webpages before training and 16 afterwards. Three experimental conditions: instructor ($n = 30$), computer (NoPhish, $n = 25$), text (NoPhish $n = 26$).	<ul style="list-style-type: none"> • Participants trained by NoPhish performed better after (81.5%) training than before (57%) training in identifying phishing websites. • This performance after training was statistically lower than people that followed classroom training (65% before training and 94% afterwards).
(Zhang-Kennedy et al., 2016)	30 university students were asked general questions about phishing before and after interacting with Geo-Phisher. Q1: what is phishing? Q2: can you describe how to protect yourself? Two treatment groups: high interactivity (Geo-Phisher, $n = 15$), low interactivity (static online training, $n = 15$).	<ul style="list-style-type: none"> • 60% of the low interactivity condition answered Q1 correctly before training, and 73.3% after training. • 80% of the high interactivity condition answered Q1 correctly before training, and 88% after training. • The difference in posttest results for Q1 was not significant. • 0% of the low interactivity condition answered Q2 correctly before training, and 26.6% after training. • 6.67% of the high interactivity condition answered Q2 correctly before training, and 73.3%% after training. • The difference in posttest results for Q2 was significant.
(Lastdrager et al., 2017)	353 children participated in an anti-phishing training experiment on the effect of classroom training (presentation and discussion) on the ability to distinguish phishing emails and websites from legitimate emails and websites. Two treatment groups: control ($n = 172$); intervention ($n = 181$)	<ul style="list-style-type: none"> • Trained children improved their ability to identify emails by 14% after training. • After four weeks the enhanced phishing detection ability of trained children returned to pre-training levels. • Over four weeks the ability to recognize legitimate emails increased.

Other articles in the Literature review.	
Author	Result
(Downs et al., 2006)	Awareness of the risk of phishing is not linked to users' feeling of vulnerability, or to strategies that help them to avoid the attacks.
(Dhamija et al., 2006)	Visual deception attacks can make sophisticated users fall for phishing.
(Robila & Ragucci, 2006)	Class discussions have a positive influence on the performance in phishing IQ tests.
(Basnet et al., 2008)	Provide cues to detect phishing emails and propose a technical countermeasure.
(Dong et al., 2008)	The s in https:// and a lock in the URL bar do not necessarily mean that a website is legitimate.
(Herzberg & Jbara, 2008)	Analyzed phishing attacks and proposed a technical countermeasure with secure identification indicators.
(Alnajim & Munro, 2009b)	Examined the retention effect of APTIPWD.
(Kumaraguru, Cranor, et al., 2009)	Developed a training intervention that replaces blacklisted websites.
(Marett & Wright, 2009)	Examined the properties in phishing emails that may influence users to give personal information.
(Arachchilage & Cole, 2011)	Described the development of anti-phishing training for the smartphone.
(Jansson & Von Solms, 2011)	Described the development of a phishing awareness training experiment. The results of the experiment were presented in a another article (Jansson & Von Solms, 2013).
(Arachchilage & Love, 2013)	Described features necessary to implement in a anti-phishing game for the smartphone.
(Kearney & Kruger, 2013)	69% of all employees that fell for phishing had followed corporate training.
(Kearney & Kruger, 2014)	92% of all employees that fell for phishing had followed corporate training.
(Aggarwal et al., 2014)	Examined 600 phishing emails without embedded links to identify its characteristics.
(Hale et al., 2015)	Proposed a new game with a realistic setting to address the benefits of embedded training and the benefits of game based training.

Other articles in the Literature review.	
Author	Result
(Parsons et al., 2015)	Informed participants that participated in this phishing study were significantly better in identifying phishing attacks than uninformed participants.
(Abbasi et al., 2016)	Users can be segmented based on their phishing susceptibility.
(Arachchilage et al., 2016)	Proposed a technical countermeasures and ways to identify phishing URLs.
(Clark & Mayer, 2016)	A book about learning principles that suggests ways to teach users in an electronic environment.
(Jeeva & Rajsingh, 2016)	Proposed technical countermeasures and ways to identify phishing URLs.
(Kritzinger, 2016)	Proposed posters as a short-term solution for phishing at schools with few resources.
(Sun & Lee, 2016)	Examined the difference between classroom training, worksheets, and drawing concept maps on the ability of users to identify phishing attacks.

Appendix 2: Methodological Design Anti-Phishing Training Studies

	<i>N</i>	Training	Real world test	Corporate setting	Control group	Enhance phish avoiding	Outperform control
(Dodge & Ferguson, 2006)	4,118	NT	√			√	N.A.
(Kumaraguru, Rhee, Acquisti, et al., 2007)	30	PG		√	√	√	√
(Kumaraguru, Rhee, Sheng, et al., 2007)	42	PG		√	√	√	√
(Sheng et al., 2007)	42	PP			√	√	√
(Kumaraguru et al., 2008)	321	PG	√	√	√	√	√
(Alnajim & Munro, 2009a)	36	WM		√	√	√	
(Kumaraguru, Cranshaw, et al., 2009)	515	PG		√	√	√	√
(Smith et al., 2009)	46	SE			√	√	√
(Aburrous et al., 2010)	100	NT			√	√	√
(Davinson & Sillence, 2010)	64	PP	√		√	√	
(Kumaraguru et al., 2010)	4,517	PP			√	√	√
(Bowen et al., 2011)	2,000	WM	√			√	N.A.
(Dodge et al., 2012)	892	TM	√		√	√	√
(Mayhorn & Nyeste, 2012)	84	PP			√	√	
(Yang et al., 2012)	62	AE			√	√	
(Jansson & Von Solms, 2013)	25,579	SW	√			√	N.A.
(Caputo et al., 2014)	1500	TT	√	√	√	√	
(Gupta & Kumaraguru, 2014)	3,359	LP	√			√	N.A.
(Zielinska et al., 2014)	96	FE			√	√	
(Canova et al., 2015)	19	NP				√	N.A.
(Arachchilage et al., 2016)	20	PG				√	N.A.
(Lim, Park, & Lee, 2016)	1045	NT	√	√		√	N.A.
(Stockhardt et al., 2016)	81	NP			√	√	

Note: NT = No Training, PG = PhishGuru, PP = Anti-Phishing Phil, WM = Warning Message, SE = Social-Ed, TM = Training Message, AE = Anti-Phishing Education Game, TM = Text Training, LP = Anti-Phishing Landing Page, FE = Fear Training, NP = NoPhish.

Appendix 3: Training Content of Anti-Phishing Training Studies

Study	Email	URL
(Dodge & Ferguson, 2006)	N.A.	N.A.
(Kumaraguru, Rhee, Acquisti, et al., 2007)	√	
(Kumaraguru, Rhee, Sheng, et al., 2007)	√	
(Sheng et al., 2007)		√
(Kumaraguru et al., 2008)	√	
(Alnajim & Munro, 2009a)		√
(Kumaraguru, Cranshaw, et al., 2009)	√	
(Smith et al., 2009)	N.A.	N.A.
(Aburrous et al., 2010)		
(Davinson & Sillence, 2010)		√
(Kumaraguru et al., 2010)		√
(Bowen et al., 2011)	√	
(Dodge et al., 2012)	√	
(Mayhorn & Nyeste, 2012)	√	√
(Sercombe & Papadaki, 2012)	N.A.	N.A.
(Yang et al., 2012)		√
(Jansson & Von Solms, 2013)	√	
(Caputo et al., 2014)	√	
(Gupta & Kumaraguru, 2014)	√	
(Zielinska et al., 2014)	√	√
(Canova et al., 2015)		√
(Arachchilage et al., 2016)		√
(Stockhardt et al., 2016)		√
This study	√	√

Note: N.A. means the study did not provide information about the training content, participants were not trained but for example shown an error message, or the effect of training was measured by general questions like "what is phishing?"

Appendix 4: Statistical Analysis**Table 5: Gender Frequencies**

Group	Male	Female
Control ($n = 18$)	61.1%	38.9%
Experimental ($n = 18$)	66.7%	33.3%
Total ($N = 36$)	63.9%	36.1%

Table 6: Age Frequencies

Group	Minimum age	Maximum age	Mean
Control ($n = 18$)	21.00	28.00	24.61
Experimental ($n = 18$)	21.00	54.00	26.33
Total ($N = 36$)	21.00	54.00	25.47

Table 7: Education Level Frequencies

Group	Practical	Bachelor	Master
Control ($n = 18$)	11.1%	50.0%	38.9%
Experimental ($n = 18$)	16.7%	55.6%	27.8%
Total ($N = 36$)	13.9%	52.8%	33.3%

Table 8: Mean Differences in Pretest Scores for Email set A, B, and C

		Sum of Squares	df	Mean Square	F	Sig.
TCR1	Between Groups	.057	2	.029	1.273	.293
	Within Groups	.742	33	.022		
	Total	.799	35			
PR1	Between Groups	.107	2	.053	.998	.379
	Within Groups	1.763	33	.053		
	Total	1.870	35			
LR1	Between Groups	.096	2	.048	1.695	.199
	Within Groups	.930	33	.028		
	Total	1.026	35			

Table 9: Correlation between Gender, Age, Education and TCR

			TCR1
Spearman's rho	Gender	Correlation Coefficient	-.248
		Sig. (2-tailed)	.095
		<i>N</i>	36
	Age	Correlation Coefficient	.330
		Sig. (2-tailed)	.053
		<i>N</i>	35
	Education	Correlation Coefficient	.150
		Sig. (2-tailed)	.375
		<i>N</i>	36

Table 10: Two-way ANOVA Interaction Effect of *time*group* on TCR

Source		Type III Sum of Squares	<i>df</i>	Mean Square	<i>F</i>	Sig.	Partial Eta Squared
Time	Sphericity Assumed	.199	2	.100	6.997	.002	.171
	Greenhouse-Geisser	.199	1.924	.103	6.997	.002	.171
	Huynh-Feldt	.199	2.000	.100	6.997	.002	.171
Time *	Sphericity Assumed	.147	2	.073	5.161	.008	.132
	Greenhouse-Geisser	.147	1.924	.076	5.161	.009	.132
	Huynh-Feldt	.147	2.000	.073	5.161	.008	.132
Error(Ti me)	Sphericity Assumed	.967	68	.014			
	Greenhouse-Geisser	.967	65.416	.015			
	Huynh-Feldt	.967	68.000	.014			

Table 11: Two-way ANOVA Interaction Effect of *time*group* on PR

Source		Type III Sum of Squares	<i>df</i>	Mean Square	<i>F</i>	Sig.	Partial Eta Squared
Time	Sphericity Assumed	.961	2	.480	14.898	.000	.305
	Greenhouse-Geisser	.961	1.861	.516	14.898	.000	.305
	Huynh-Feldt	.961	2.000	.480	14.898	.000	.305
Time *	Sphericity Assumed	.740	2	.370	11.475	.000	.252
	Greenhouse-Geisser	.740	1.861	.398	11.475	.000	.252
	Huynh-Feldt	.740	2.000	.370	11.475	.000	.252
Error(Ti me)	Sphericity Assumed	2.193	68	.032			
	Greenhouse-Geisser	2.193	63.279	.035			
	Huynh-Feldt	2.193	68.000	.032			

Table 12: Two-way ANOVA Interaction Effect of *time*group* on LR

Source		Type III Sum of Squares	<i>df</i>	Mean Square	<i>F</i>	Sig.	Partial Eta Squared
Time	Sphericity Assumed	.009	2	.004	.169	.844	.005
	Greenhouse-Geisser	.009	1.974	.005	.169	.842	.005
	Huynh-Feldt	.009	2.000	.004	.169	.844	.005
Time *	Sphericity Assumed	.021	2	.010	.395	.675	.011
	Greenhouse-Geisser	.021	1.974	.011	.395	.672	.011
	Huynh-Feldt	.021	2.000	.010	.395	.675	.011
Error(Ti me)	Sphericity Assumed	1.784	68	.026			
	Greenhouse-Geisser	1.784	67.102	.027			
	Huynh-Feldt	1.784	68.000	.026			

Table 13: Two-way ANOVA Interaction Effect of *time*group* on Confidence

Source		Type III Sum of Squares	<i>df</i>	Mean Square	<i>F</i>	Sig.	Partial Eta Squared
Time	Sphericity Assumed	.282	2	.141	1.187	.311	.034
	Greenhouse-Geisser	.282	1.667	.169	1.187	.306	.034
	Huynh-Feldt	.282	1.795	.157	1.187	.308	.034
Time *	Sphericity Assumed	.718	2	.359	3.019	.055	.082
	Greenhouse-Geisser	.718	1.667	.431	3.019	.066	.082
	Huynh-Feldt	.718	1.795	.400	3.019	.062	.082
Error(Ti me)	Sphericity Assumed	8.086	68	.119			
	Greenhouse-Geisser	8.086	56.690	.143			
	Huynh-Feldt	8.086	61.017	.133			

Note: In this analyses the Greenhouse-Geisser method is used because the assumption of sphericity was violated according to Mauchly's test of sphericity : $\chi^2(2) = 7.343, p = .03$