Heliyon 5 (2019) e02010

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.heliyon.com

Effectiveness of and user preferences for security awareness training methodologies

Kai Florian Tschakert *, Sudsanguan Ngamsuriyaroj

Faculty of Information and Communication Technology, Mahidol University, Thailand

ARTICLE INFO

Keywords:
Computer science
Education
Cyber attack
Computer fraud
Computer security training
Information security
Phishing
Security awareness training
Cyber security
Information security

ABSTRACT

Phishing is a primary vector used in cyber-attacks, and current technical measures are not sufficient to reduce their success to an acceptable level. Empowering users to identify phishing emails is crucial; thus, anti-phishing training is essential. We investigate participant phishing susceptibility in a 2 × 2 mixed factorial design to determine if instructor-led classroom training, in addition to a multiple approach video-, game-, and text-based training package, offers a significant difference in susceptibility reduction compared with the absence of classroom training. The results suggest an insignificant improvement in reducing phishing susceptibility by incorporating classroom training. Furthermore, we observe a significant preference from the participants for one training method (i.e., classroom training) only if a decision for one particular method was required.

1. Introduction

Phishing is a form of social engineering observed since the mid-1990s and remains a problem even today [1]. In 2016, the incident response firm Mandiant [2] reported that spear phishing constitutes 25% of the attack vectors observed in EMEA[1], making it the second-most used attack vector. Publications from other security service providers, including Kaspersky [3], Verizon [4], and CyberArk [5], suggest a similar result. Europol [6] also noticed an "increase of targeted phishing aimed at high-value targets" [6] as well as an increase in "the overall quality and authenticity of phishing campaigns" [6].

Studies examining the use of security education and awareness training to reduce users' susceptibility to phishing attempts demonstrate that education is a successful approach measurable by the false-negative rate [7, 8, 9, 10, 11, 12, 13, 14]. Furthermore, Stockhardt et al. [15] and Abawajy and Kim [16] showed that distinct types of training have different impacts on the ability to identify phishing emails. Additional experts [17] agree that a mix of training approaches is most promising. While many studies about phishing exist, more research is required as the problem remains prevalent and is expected to continue to be important with the increasing use of electronic communication. We found no research conducted in Thailand that focused on training to reduce phishing susceptibility, and of the studies reviewed, most used only questionnaires (i.e., screenshot-based) or in-lab tests instead of simulated phishing emails. The supplemental material "list of relevant previous literature" gives an overview of previous literature relevant to our study, none of them with a Thai population.

In this study, we compare the impact of training on two groups of participants (referred to as Group A and Group B) who received a combination of text-based, video-based, and game-based education about phishing. Group B additionally received classroom-based instructor-led training. Our objective is to investigate the effectiveness of the training combinations. The impact of the training is measured by sending imitation phishing emails to the participants before and after the training, followed by a comparison of the false-negative rate (i.e., measuring how many participants clicked on the phishing links and then provided data submission through the subsequent phishing page). Additionally, a screenshot-based questionnaire test was conducted to measure a possible change in the false-negative and false-positive rates (i.e., measuring how many participants rated a legitimate website or email as phishing), which indicates that participants are more alert but no better in distinguishing phishing and legitimate emails. The participants also received a questionnaire before and after the training to measure their opinions about the process and whether their knowledge increased. The post-training

* Corresponding author.
E-mail address: florian.tschakert@gmail.com (K.F. Tschakert).
[1] Europe, the Middle East and Africa.

https://doi.org/10.1016/j.heliyon.2019.e02010
Received 19 February 2019; Received in revised form 24 May 2019; Accepted 11 June 2019

questionnaire is partly based on Stockhardt et al. [15].

Our results show that the overall training has a significant positive impact on reducing phishing susceptibility but that no single training combination is best. They also show that participants do not have a preference for one training combination but do prefer the classroom-based instructor-led training if it is included in the approach they experienced.

2. Materials and methods

2.1. Training formats and materials

The selection of the materials focuses on reducing the time participants invest in the study and not requiring deep technological understanding. The materials should be easy to access as recommended in [16] and should be freely available or straightforward to recreate so that the approach may be used by organizations with a limited budget. All training material was developed in the English language as the participants were recruited from faculty with English as their primary communication and instruction language. Based on these considerations, we used the following training materials.

- *Video-based training (self-paced, both groups):* YouTube features a variety of videos available for our purpose, but there exists no measure for video quality. However, for this study, the videos considered are not specific to a company or organization and are not part of an advertisement campaign by a vendor. An additional requirement is that the video must focus on phishing and should provide actionable information on how to detect phishing. The selected videos are short (approximately two to 3 min) and presented with clear and understandable English language. We identified two videos, "*What is Phishing?*" (3:08 min) [18] and "*What is Phishing and How do I Protect Myself*" (2:28 min) [19], to be suitable for selection. The first provides a general introduction to phishing, a brief overview of the techniques used by cybercriminals, their potential consequences, and the indicators to help detect illegitimate emails and URLs. The second provides additional instruction on how to avoid being deceived by phishing emails or URLs.
- *Game-based training (self-paced, both groups):* Many studies [10, 11, 13, 15, 16, 20, 21, 22] leverage game-play to interactively educate users. Training games allow users to play interactively with challenges to decide if emails are trustworthy. Knowledge transfer about phishing methodologies and how to detect phishing emails is typically incorporated into the game. This study uses the web browser-based game *Anti-Phishing Phil* [13] (freely accessible at http://www.ucl.ac.uk/cert/antiphishing/) and *Anti-Phishing Phyllis (*free demo version at https://beta.wombatsecurity.com/webdemo/4.7/?module=phyllis). Both games played completely take between five and 10 min.
- *Text-based training (self-paced, both groups):* An educational text was created based on freely available material [13, 15, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31] including examples and describing the best practices to detect phishing emails. The content is identical to the instructor-led training presentation with a brief introduction to phishing, an example with its possible consequences, indicators of illegitimate emails and URL addresses, and additional examples of phishing emails and websites. We expect a participant spends 15–20 min reading the text.
- *Instructor-led classroom (scheduled, Group B):* Conducting anti-phishing training through a lecture setting with an instructor has only been done previously by Stockhardt et al. [15] and exhibited a significant positive impact on the participants' ability to detect phishing emails. As the lecture is provided at a specified time requiring attendance, this approach is not self-paced, unlike the prior methods. As stated by Kumaraguru et al. [9], this method is "time--consuming for employees and expensive for companies." Therefore, it is essential to understand the justification for this effort. Our instructor-led training is the experimental distinction between our study groups A and B. The content created for this study is based on the same free material provided in the self-study text brochure, making it similar to the text-based approach with the examples presented interactively through an in-class quiz. Here, participants are asked to raise their hand if they believe the example represents a phishing or a legitimate email or website. Participants are also encouraged to ask questions during the presentation or the concluding question and answer (Q&A) session. The classroom training lasted 30 min with an additional 15 min for the Q&A session.

The three self-paced training methods were provided to the participants via email. It was not determined if the participants completed all the material but only if they opened the material. These materials required 25–35 min to complete while the scheduled classroom-based training lasted an additional 30–45 min. Thus, if consuming all training material, Group B participants spent more than twice the time to complete the training.

2.2. Participant group and recruitment

Voluntary participants were recruited among the fourth-year computer science faculty of Mahidol University's Information and Communication Technology department. This group was comprised of a homogenous group of participants with experience in being instructed in English. The research project received the approval of Mahidol University's Institutional Review Board (IRB), and the recruitment process followed their requirements. The consent form described the research on Phishing and Awareness Training, but the participants were not informed that contrived phishing emails would be a part of the study. This approach reduces the bias (alerted state) of the participants, as argued in [32] and [33], and was previously proven to influence results by [25]. Initially, we recruited 50 participants, but 17 did not complete the project to the end. Therefore, we had 17 participants in Group A and 16 in Group B. Two participants from Group B did not join the classroom training; thus, they were excluded in much of the statistical analyses. If they were included, then we labeled the data as being from Group B-Plus. Table 1 provides the demographics of our participants.

We asked for the participants' age in ranges of 18–19, 20–21, and 22–23 years. Group A had 13 (76.5%) participants answer as 20–21 years and four (23.5%) as 22–23 years. In Group B, one individual (6.3%) answered as 18–19 years, eight participants (50%) answered 20–21, and six (37.5%) answered 22–23, while one (6.3%) did not answer.

We chose a homogenous group to reduce potential demographic effects. Previous studies have come to different conclusions of whether there is an effect of age. Several studies [7, 8, 13, 34, 35] did not find an effect of age, while other studies [11, 36, 37] found that younger participants were more susceptible to phishing attacks. Among university students, Mohebzada et al. [38] found senior students most susceptible. Thus, if there is an effect of age, we chose the group being most susceptible to phishing and therefore in most need of receiving training.

Participants were also asked to provide their most commonly used email addresses, which resulted in 20 (60.6%) Gmail addresses, eight (24.2%) university email addresses, and five (15.2%) Hotmail/Live.com addresses. This distribution indicates that most of the participants provided their frequently used email address and not merely their university

Table 1
Participant demographics.

Gender	Group A (N = 17)	Group B (N = 16)	Group B-Plus (N = 18)
Male	7 (41.2%)	8 (50%)	8 (44,4%)
Female	10 (58.8%)	7 (43.5%)	9 (50%)
Did not state	0	1 (6.3%)	1 (5,6%)
Median age	20–21	20–21	20–21

addresses. If only the university email addresses were provided, then the simulated phishing emails would, in many cases, be sent to accounts not often used, so may not be readily noticed at all.

2.3. Research questions

The following research questions driving this study are in part based on [15].

Question #1 – Training method combination effectiveness: (a) Does the overall training plan significantly decrease phishing susceptibility? (b) Is there a significant difference in effectiveness between the training combinations?

Question #2 – Participant confidence as in [15]: (a) Does the user self-assessment increase due to the training? (b) Is there a significant difference in the users' self-assessment between the training combinations?

Question #3 – Training satisfaction as in [15]: Is there a significant difference in satisfaction between the training combinations?

Question #4 – Training method preference: (a) Is one training method within each combination considered more helpful compared to the others? (b) Is one training method within each combination preferred by participants compared to the others?

Stockhardt et al. [15] compared text-, game-, and instructor-led classroom-training with participants from a vocational college in Germany. The classroom-based training achieved the best result in reducing phishing susceptibility, significantly better than game-based training, but with an insignificant difference from text-based training. The overall satisfaction of participants is not significantly different among the three training methods. All training methods resulted in a significant increase of the participants' confidence in their ability to detect phishing. Because an objective of our study was to evaluate the benefit of classroom-training and the user preference for training methods, the questions designed by Stockhardt et al. [15] are suitable for our study. Furthermore, using the same questions allows us to compare our results with the Stockhardt results. Furthermore, the demographic features of both studies are comparable with the exception of theirs being conducted with German students and ours with Thai students.

2.4. Measures

The measures are based on Signal Detection Theory (SDT) [39], which can be used, as shown by Sheng et al. [13], in the context of information security awareness training to determine if training increases alertness or the ability to detect phishing. Using SDT, we can "quantify the ability to discern between signals (phishing websites in this case) and noise (legitimate websites in this case)" [13]. SDT measures include hits, misses, false alarms, and correct rejections. In the context of this study, a hit represents a correctly identified phishing website or email, a miss is a phishing website or email identified as legitimate, a false alarm is a legitimate website or email identified as phishing, and a correct rejection is a correctly identified legitimate website or email (see Fig. 1). Thus, a measure of the false-negative rate is represented as the fraction of misses for all signals (phishing emails and websites), and the measure of the false positive rate is the fraction of false alarms for all noise (legitimate emails and websites).

We consider a security awareness training event as successful for reducing phishing susceptibility within a group if the false-negative rate reduces following training. However, if the false-positive rate increases simultaneously, then this is an indicator that the participant is more alerted and has a greater tendency to rate both illegitimate and legitimate websites or emails as phishing. Thus, the training would be considered as not successful for teaching users how to detect phishing.

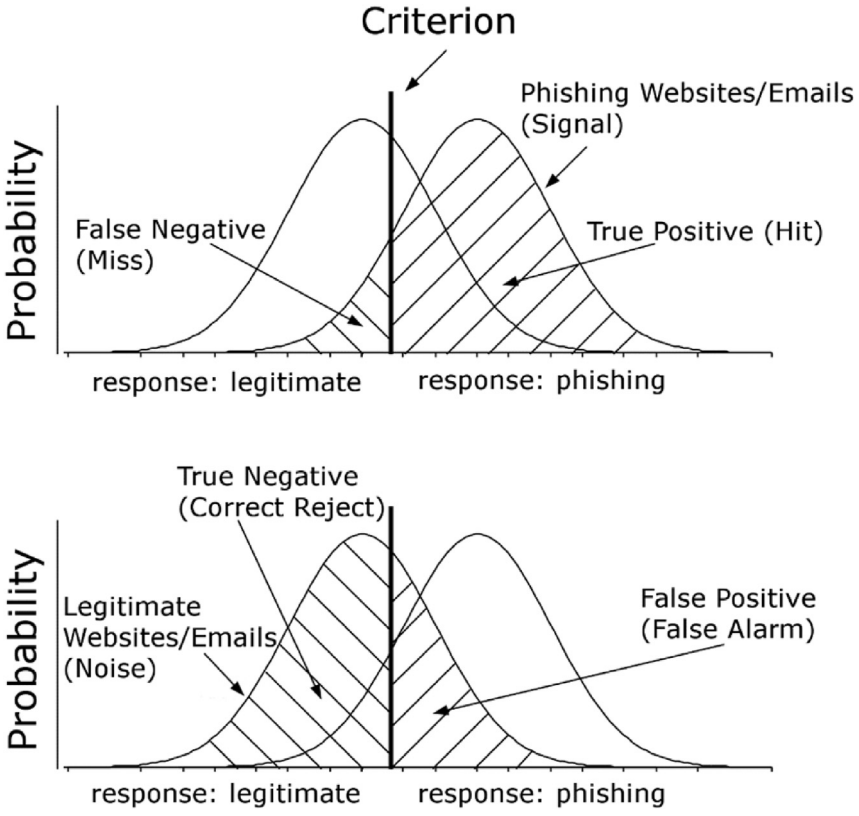

Fig. 1. Illustration of SDT as used in our study. Criterion is the decision tendency. A training method that we consider as successful would reduce the false-negatives without increasing the false-positives. This means, we do not want to see the participant to just shift their criterion to the left (i.e., becoming more alerted) but rather to separate both distributions better. The figure is based on Heeger [40] and Sheng et al. [13].

2.5. Research design

Using a "within-between subject design" (i.e., a split-plot design) [41], we answered Question #1 with a pre- and post-training questionnaire as well as simulated phishing emails sent before and after the training session. Question #2 was answered with these same pre- and post-training questionnaires. Questions #3 and #4 were answered based only on the post-questionnaire using a between-subject design.

The study is divided into the phases of "pre-training simulated phishing," "pre-questionnaire," "training session," and "post-training simulated phishing and post-questionnaire." We distributed the pre-test simulated phishing emails before providing the pre-questionnaire as it made the participants indirectly aware of the phishing attempt.

- *Pre-training simulated phishing*: Parsons et al. [25] used 25 phishing emails as stimuli in their study and selected each based on the Global Phishing Survey [42], stimuli used in previous research, and what they would expect in a student's inbox. Each stimulus used by Parsons et al. was an email the authors received or found online. They categorized the emails into four intention-based collections of "benefit or gain" (i.e., a call to action, such as clicking a link or providing information to win a prize), "risk or loss" (i.e., a call to action, such as clicking a link or providing information to avoid a loss), "account information" (i.e., information related to an account with no call to action), and "information only" (i.e., information unrelated to an account and no call to action).
  We sent four simulated phishing emails to each participant before the training event. Each email represented one of the intention-based categories because we determined they matched well with our experience and they offered a diverse set of stimuli so that the probability that participants would relate the email to this experiment was low (e.g., due to identifying these emails as part of our testing). The first email featured the "benefit or gain" template (1) with an opportunity to win an iPhone 8. The second email was a "risk or loss" (2) that alerted the participant to reconfirm their university e-learning system account. The third used the "account information" email (3) to inform the recipient that someone shared a file with them. The fourth email was "information only" (4) to lure the participant with gossip about a superstar. Data entry was possible in the first three emails, and we measured for each email how many participants clicked the included link at least once and submitted data at least once.

- *Pre-questionnaire*: A pre-training questionnaire collected the participants' demographic information as well as a ranking of the statement, "I know how to protect myself from phishing" to provide a self-assessment. This input selection included the five options of "strongly agree" (5 points), "agree" (4), "undecided" (3), "disagree" (2), and "strongly disagree" (1). Next, we asked the participants to review 20 screenshots and decide if each represented a phishing or legitimate email or website. The collection included five screenshots each of legitimate and phishing emails and websites. The websites and emails presented were real in appearance, but with modified URLs from altering the source code locally in the web browser for each email and the browser's address bar. We leveraged the seven deception categories, based on examples taken from PhishTank[2] and defined by Canova et al. [22], to create the phishing URLs. These categories are based on difficulty and do not consider methods impossible to detect by eye (i.e., an internationalized domain name used in homograph attacks) or not suitable for the use on screenshots (i.e., redirections). Moreover, we did not make use of the two easiest categories. Table 2 includes additional details about the screenshots presented in the questionnaire and Fig. 2 is an exemplar stimulus used in the questionnaires. The stimuli were not randomized in their order as we did not assess the difficulty of the stimuli but rather the ability of the participants to assess them. Therefore, we decided against randomization to avoid differing order effects within and between the relatively small groups.

- *Training session*: The training methods were delivered by email with Group B additionally receiving the classroom-based event offered on three dates. The participants had eight days to work on the material which was then extended by one day following a reminder email.
- *Post-training simulated phishing and post-questionnaire*: We sent four emails similar to the pre-training emails with alternate contexts. Here, the messages claimed that the participants (1) could win an iWatch-3 LTE, (2) should extend their university Wi-Fi account to avoid cancellation (see Fig. 3), (3) someone shared a song with them through a streaming platform, or (4) are informed about surprising news. The questionnaire contained the same self-assessment as included in the pre-questionnaire along with a rating of the participant's satisfaction with the questions, based on [15], including (a) "I enjoyed learning about phishing the way I did," (b) "I think I learned a lot," and (c) "What I learned helped me to protect myself from phishing attacks." Additionally, the participants rated the statement, "the training method helped me protect myself from phishing attacks" based on the training method included in combination with the final question of "which training method did you like the most?" Finally, the participants again assessed the same screenshots from the pre-questionnaire.

The schedule for these four phases followed:

- *Pre-training simulated phishing*: Day 1 (Friday) through day 8 (Friday). The simulated phishing emails were distributed throughout the eight days to avoid alerting participants by receiving such emails at a higher-than-typical frequency.
- *Pre-questionnaire*: Day 9 through day 17 (Saturday to Sunday) followed by two break days.

Table 2
Screenshots for assessment presented in the questionnaires within the deception category.

ID	Description	Type	URL Deception Category based on [22]
Emails			
1	Banking	Phish	"Random/Unrelated/Trustworthy/IP Domain, with Brand in Path"
2	University	Phish	"Derived Domains" (modified top-level domain)
3	Airline	Legit	n/a
4	University	Legit	n/a
5	Banking	Legit	n/a
6	Airline	Phish	"Random/Unrelated/Trustworthy Domain, with Brand in Subdomain"
7	E-Commerce	Phish	"Derived Domains"
8	E-Commerce	Legit	n/a
9	Social Media	Phish	"Introducing Typos" plus modified top-level Domain
10	E-Commerce	Legit	n/a
Websites			
11	Social Media	Phish	"Random/Unrelated/Trustworthy/IP Domain, with Brand in Path"
12	E-Commerce	Legit	n/a
13	Banking	Phish	"Derived Domains"
14	Banking	Legit	n/a
15	Social Media	Legit	n/a
16	E-Commerce	Phish	"Replacing Character(s)"
17	University	Legit	n/a
18	Network Provider	Legit	n/a
19	Network Provider	Phish	"Random/Unrelated/Trustworthy Domain, with Brand in Subdomain"
20	University	Phish	"Derived Domains" (modified top-level domain)

[2] https://www.phishtank.com.

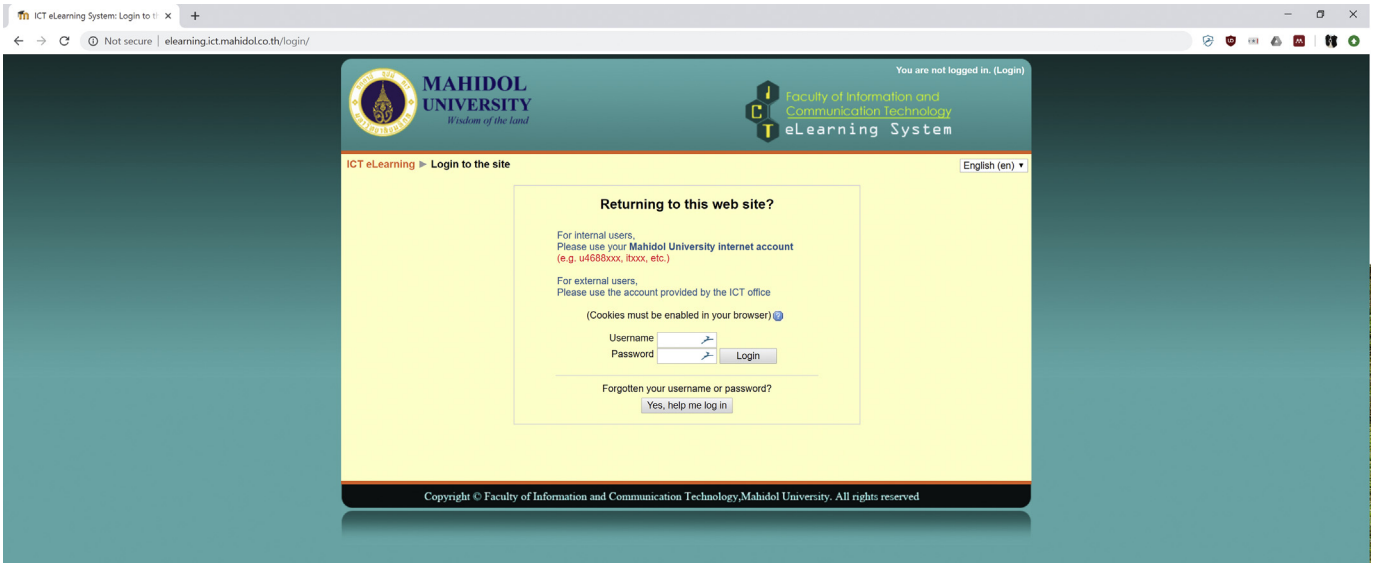

Fig. 2. Questionnaire stimulus ID20 from Table 2 showing a phishing website where the domain mahidol.co.th is used instead of the real domain mahidol.ac.th.

- *Training session*: Day 20 (Wednesday) through day 28 (Thursday) with classroom training provided during two sessions on day 26 and one session on day 27. The participants were reminded to complete the self-paced training and to participate in the classroom training on day 25 during a lecture and via email on day 27.
- *Post-training simulated phishing and post-questionnaire*: Day 29 (Friday) through day 36 (Friday). The simulated phishing emails were distributed throughout the eight days as was done during the pre-training simulated phishing process.

3. Results

The evaluation results reviewed below are structured according to the four research questions. The first section provides results for the training combination effectiveness, the second section covers the results for participant confidence, the third includes the participants' satisfaction with the training, and the fourth section reviews the results of the participants' preferences for each training method offered.

The analysis of the results shows that the participants downloaded the self-paced training material with 13 of 17 from Group A opening the materials and 12 of 16 from Group B. As the participants know each other, it is possible they did not open their personalized links and instead consumed the training material together. With the post-questionnaire, they also had the opportunity to answer for each training method they did not experience. In Group B, only one of the 16 participants answered that they did not use the video-based and text-based training, so we assume that the participants consumed the training materials as there was no incentive to lie.

3.1. Question #1: training combination effectiveness

The experimental design features a 2 × 2 mixed factorial design as described in [43] with a test-time (pre, post) x type of training (classroom, without classroom) matrix of factors. The time varies within the subject while the training varies between subjects, as in [10]. The analysis was conducted based on the simulated phishing emails and the questionnaire.

Before the training, 13 (9.3%) links in the simulated phishing emails were clicked, and 127 (90.7%) were not clicked. For 12 out of the 13 clicked emails, there was an option provided to submit data. In 10 (83.3%) out of these 12, participants submitted data. After the training, only two (1.4%) links in the phishing emails were clicked and 138 (98.6%) were not clicked. For both clicked emails, the option to submit data was provided. In two (100%) of these cases, the participant submitted the data.

Before the training, Group A participants clicked nine (13.2%) phishing links and Group B clicked two (3.1%) phishing links. Group A submitted data in seven (87.5%) of eight possible cases, and Group B participants submitted data in one (50%) of two cases. After the training for both groups, only one phishing link was clicked per group representing that Group A clicked 1.5% and Group B clicked 1.6% of the phishing links. For both groups, data were submitted after clicking the link (100%).

Owing to the small number of clicks and large difference between the two groups, the requirements for the intended statistical analysis (Split Plot ANOVA [41, 44]) were violated; thus, analysis was not possible. Therefore, we transformed the data to enable a conclusion. The false-negative rate was transformed into a dichotomous variable set as zero if no phishing email was clicked and one if any phishing email was clicked. We included all 35 participants from Group A and Group B-Plus. Before the training session, 12 (34.3%) participants were phished, and 23 (65.7%) were not phished. Following the training session, the number of non-phished participants increased to 33 (94.3%) with an associated reduction of phished participants to only two (5.7%).

We ran an exact McNemar's test [45] to determine if there was a difference in the proportion of non-phished participants compared with pre- and post-training. The proportion of non-phished participants increased from 0.66 to 0.94 with a statistically significant difference of $p = 0.006$. Overall, the reduction of participants who succumbed to the simulated phishing was significant. However, this data does not allow for an answer as to whether there is a difference between the training combinations.

The screenshot assessment in the questionnaires was used in comparison to the simulated phishing emails to answer Question #1. Before the training, 239 (68.3%) emails and 252 (72%) websites were assessed correctly and 111 (31.7%) and 98 (28%) falsely, respectively. After the training, 249 (71.1%) emails and 286 (81.7%) websites were assessed correctly and 101 (28.9%) and 64 (18.3%) falsely, respectively. The correct decisions for all scenarios increased from 491 (70.1%) to 535 (76.4%) and wrong decisions decreased from 209 (29.9%) to 165 (23.6%).

Group A increased their correct overall decisions from 251 (73.8%) to 263 (77.4%) after the training, while decreasing wrong decisions from 89 (26.2%) to 77 (22.6%). Group B (excluding the participants who did not

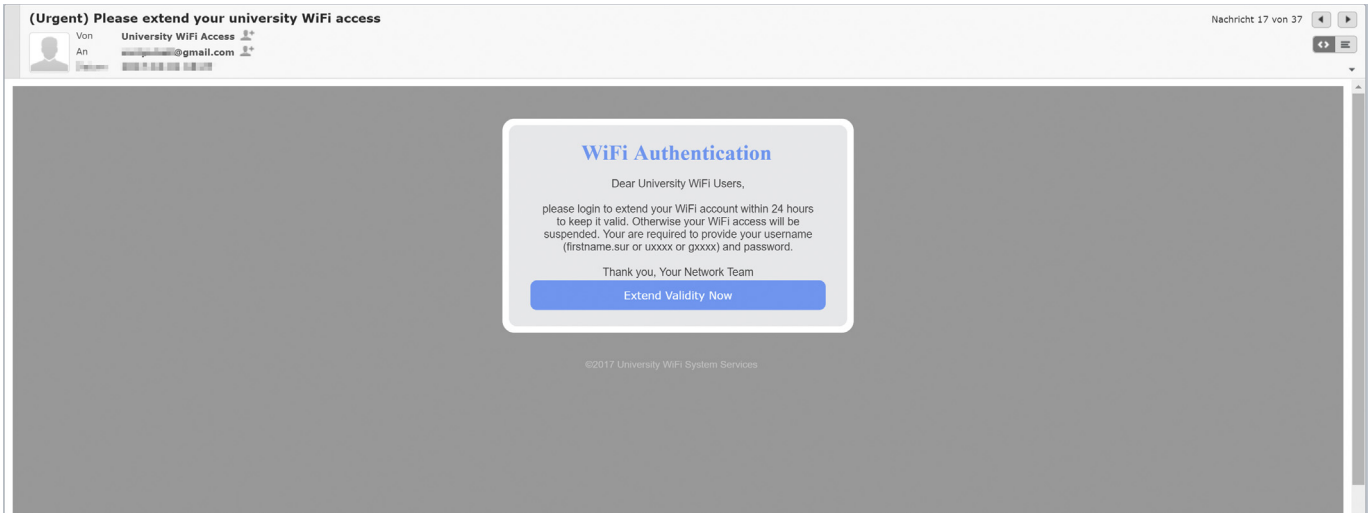

Fig. 3. Stimuli from the post-training phishing emails sent to the participants. The "risk or loss" email alerted the participants to log into their university Wi-Fi account to avoid being suspended.

participate in the classroom training) increased correct decisions from 215 (67.2%) to 249 (77.8%) after the training, while decreasing wrong decisions from 105 (32.8%) to 71 (22.2%).

We ran a split-plot ANOVA with the group as the between-subjects independent variable, time (pre- or post-training) as the within-subjects independent variable, and the false-negative rate of the screenshot assessment as the dependent variable. No statistically significant interaction was measured between group and time on the screenshot assessment false negative rate, $F(1, 31) = 2.705$, $p = 0.11$, and partial $\eta 2 = 0.080$. The main effect of time showed a statistically significant difference in the screenshot assessment false-negative rate at the different time points, $F\ (1, 31) = 4.647$, $p = 0.039$, and partial $\eta 2 = 0.130$, and the false-negative rate was significantly lower post-training ($M = 0.074$, $SE = 0.035$, $p = 0.039$). As no opportunity for interaction was offered, a difference between the groups was not measured, but the overall training was observed to decrease the false-negative rate.

Repeating this analysis with the false-positive rate of the screenshot assessment as the dependent variable, we again found no statistically significant interaction between group and time, $F\ (1, 31) = 0.254$, $p = 0.618$, and partial $\eta 2 = 0.008$. The main effect of time showed a statistically significant difference in the screenshot assessment false-positive rate at the different time points, $F\ (1, 31) = 5.703$, $p = 0.023$, and partial $\eta 2 = 0.155$. The false-positive rate was significantly lower post-training ($M = 0.067$, $SE = 0.028$, $p = 0.023$), and with no significant interaction, a difference between the groups was not measured, but overall the training decreased the false-positive rate.

3.2. Question #2: participant confidence as in [15]

The mean of the self-assessment score across participants increased from pre-training ($M = 3.91$, $SD = 0.77$) to post-training ($M = 4.42$, $SD = 0.56$). Group A changed from $M = 4$ ($SD = 0.71$) to $M = 4.35$ ($SD = 0.61$) and Group B from $M = 3.81$ ($SD = 0.83$) to $M = 4.5$ ($SD = 0.52$).

We ran a split-plot ANOVA with group as the between-subjects independent variable, time (pre- or post-training) as within-subjects independent variable, and the self-assessment score as the dependent variable. No statistically significant interaction between group and time was measured on the self-assessment score, $F\ (1, 31) = 1.651$, $p = 0.208$, and partial $\eta 2 = .051$. The effect of the variable time showed a statistically significant difference in the self-assessment score at the different time points, $F\ (1, 31) = 15.970$, $p < 0.001$, and partial $\eta 2 = .340$. As there was no significant interaction, a difference between the groups was not measured, but the overall training increased the self-assessment score.

3.3. Question #3: training satisfaction as in [15]

Based on the post-training questionnaire, we compared the groups' training experiences. The enjoyment and learning experience scores were higher for Group B compared to Group A, while Group A reported a slightly higher protection experience score (see Table 3).

Furthermore, the combination of the results of the three questions was calculated as an overall experience score, which resulted in a value for Group B ($M = 4.33$, $SD = 0.37$) higher than for Group A ($M = 4.12$, $SD = 0.53$).

First, we ran a Mann-Whitney U test [46] to determine differences in the overall experience score between the two groups. The distributions were similar as assessed by visual inspection. The overall experience score was not statistically significantly different, with $U = 165$, $z = 1.077$, and $p = .309$, by using an exact sampling distribution for U [47]. The overall experience score for Group B was higher without a significant difference.

Second, we ran the same test to determine differences in the enjoyment score between the two groups. The enjoyment score was not statistically significantly different, with $U = 158.5$, $z = 0.936$, and $p = 0.423$. The enjoyment score for Group B was higher without a significant difference.

Third, the learning score was not statistically significantly different with $U = 181.5$, $z = 1.816$, and $p = 0.102$. The learning score for Group B was higher without a significant difference.

Table 3
Descriptive data for the training experiences.

	Enjoyment		Learning		Protection	
	A	B	A	B	A	B
Strongly Disagree						
Disagree			1			
Undecided	2	1	4	1	1	
Agree	11	9	9	9	8	10
Strongly Agree	4	6	3	6	8	6
Mean	4.12 (SD = 0.6)	4.31 (SD = 0.6)	3.82 (SD = 0.81)	4.31 (SD = 0.6)	4.41 (SD = 0.62)	4.38 (SD = 0.5)

Fourth, the protection score was not statistically significantly different with $U = 128$, $z = -0.33$, and $p = 0.79$. The protection score for Group A was slightly higher without a significant difference.

3.4. Question #4: training method preference

Group A rated the helpfulness of the video training as the best offering over the text and game training methods, with the latter receiving identical results (see Table 4). For deciding on the most-liked method, Group A responded with video training eight times, game training seven times, and text training twice.

We ran a Kruskal-Wallis H test [48] to determine differences in the helpfulness scores between different training methods for Group A for text training (n = 17), video training (n = 17), and game training (n = 17). The distributions of the helpfulness scores were generally similar for all methods as assessed by visual inspection of a boxplot. The median helpfulness scores were 4.0 for all methods, and there were no statistically significant differences, with $\chi 2$ (2) = 1.334 and $p = 0.513$. Group A participants did not have a significant preference for the helpfulness of any specific training method.

We conducted a chi-square goodness-of-fit test [49] to determine if an equal number of participants selected each of the three training methods as their most-liked option. The minimum expected frequency was five. The chi-square goodness-of-fit test indicated three training methods were equally represented by the participants of Group A ($\chi 2$ (2) = 3.647, p = 0.161), and there was no significantly most-liked training method within Group A.

Group B rated the helpfulness of the classroom training as the best, followed by video, game, and text training, respectively (see Table 5). For deciding on the most-liked method, Group B responded with classroom training nine times, game training four times, video training twice, and text training once.

Another Kruskal-Wallis H test determined differences in the helpfulness scores between the different training methods for Group B with text training (n = 15), video training (n = 15), game training (n = 16), and classroom training (n = 16). The distributions of the helpfulness scores were mostly similar for all methods as assessed by visual inspection of a boxplot. The median helpfulness scores were 4.0 for text, video, and game training and 5.0 for classroom training, and no statistically significant difference exists with $\chi 2$ (3) = 3.702 and $p = .295$. Although the classroom training was rated better than the other methods, the Group B participants did not have a significant preference for the helpfulness of any one training method.

Finally, we conducted an exact chi-square goodness-of-fit test to determine if an equal number of participants selected each of the three training methods as their most-liked option. The minimum expected frequency was four, and the exact chi-square goodness-of-fit test indicated the three training methods were not equally represented by the participants of Group A ($\chi 2$ (2) = 9.5, p = 0.024) with just over half selecting the classroom training. As the chi-square goodness-of-fit test is considered unreliable for expected frequencies of less than five [49], the exact version of the test was utilized here.

Additionally, a Monte Carlo simulation [50] with $B = 100{,}000$ was repeated 10 times in the R environment [51] resulting in p values between $p = 0.023$ and $p = 0.025$. A post-hoc manual pairwise test was conducted with the exact binomial test [52], which indicated the proportion of text training ($p = 0.143$), video training ($p = 0.387$), and game training ($p = 1.00$) were not statistically significantly different from the expected proportion. However, the proportion of classroom training was statistically significantly higher than the expected proportion with $p = 0.007$ and 95% CI [0.30 to 0.80]. Therefore, we observed a statistically significant preference for the classroom-based training within Group B.

Table 4
Helpfulness rating for the training methods according to Group A.

	Text	Video	Game
Strongly Disagree			1
Disagree			
Undecided	5	2	1
Agree	6	7	10
Strongly Agree	6	8	5
Mean	4.06	4.353	4.06
	$SD = 0.827$	$SD = 0.702$	$SD = 0.966$

Table 5
Helpfulness rating for the training methods according to Group B.

	Text	Video	Game	Classroom
Strongly Disagree				
Disagree				
Undecided	2	1	1	1
Agree	9	9	11	6
Strongly Agree	4	5	4	9
Mean	4.13	4.267	4.19	4.5
	$SD = 0.640$	$SD = 0.594$	$SD = 0.544$	$SD = 0.632$

3.5. Analysis of statistical power

As several statistical tests did not result in significant results, we used G*Power (version 3.1.9.4) [46] to run post-hoc power analyses to determine whether the non-significances were due to low statistical power. Table 6 presents the results of the power analyses. The required sample sizes were calculated to achieve a power (1 - β) of 0.8 with $\alpha = 0.05$. The tests in rows 1, 6, and 8 would need a reasonably increased number of participants and it is likely that the statistical tests could then detect significant results. The sample size of test three would have to increase to N = 150, which is large for such a study design but not unrealistic, although 75 participants would have to participate in a classroom training. Therefore, there is a possibility that the statistical test could provide a significant result with a larger sample size. For the tests in rows 2, 4, 5, and 7 the required sample sizes are unrealistically large for such a study design. Thus, it is unlikely that the statistical tests would detect significant results with reasonably larger sample sizes.

4. Discussion

This section discusses our results and their relation to prior research categorized by the research questions.

4.1. Training combination effectiveness

The training experiences effectively and significantly decreased the phishing susceptibility of the participants. However, there was no significant difference between the two training combinations, even though the Group B participants were offered an additional training method and the total duration of their training was nearly twice that of Group A. Generally, the phishing training proved beneficial for these Thai students (with an IT background).

Results from prior studies were repeated with our demographically different group with a comparable decrease of the false-negative rates. Our study exhibited better results for the false-positive rate, which did not increase, contrary to Sheng et al. in [13] and [11] and Kumaraguru et al. [35] who observed increases. Our decreased false-positive rate indicates that the participants improved their ability to detect phishing emails and did not just become more alert, which would lead to also assessing more legitimate emails incorrectly as phishing attempts.

Comparing to Kumaraguru et al. [35], their study achieved a changed false-negative rate with a larger difference between pre- and post-training. Their study included different participant demographics, a combination of text-based and video-based training, a setup with websites presented on a local computer instead of screenshots, and a post-training test administered directly after the session. A possible

Table 6
Results of power analyses for statistical tests with non-significant results.

#	Statistical test	Effect size	Statistical power	Required sample size
Research Question #1				
1	Screenshot assessment false negative rate (ANOVA interaction effect)	$f(U) = 0.29$	0.36	96
2	Screenshot assessment false positive rate (ANOVA interaction effect)	$f(U) = 0.09$	0.08	978
Research Question #2				
3	Participant confidence rate (ANOVA interaction effect)	$f(U) = 0.23$	0.24	150
Research Question #3				
4	Overall experience score (Mann-Whitney U test)	$d = 0.46$	0.24	260
5	Enjoyment score (Mann-Whitney U test)	$d = 0.32$	0.14	330
6	Learning score (Mann-Whitney U test)	$d = 0.68$	0.46	72
7	Protection score (Mann-Whitney U test)	$d = 0.05$	0.05	11,590
Research Question #4				
8	Group A most-liked training method (chi-square goodness-of-fit test)	$w = 0.46$	0.38	46

reason for their better result may be that the post-training test was conducted immediately after the training; thus, the knowledge gained was fresh. Another reason may be that, as the false-positive rate increased, the users became more alert.

Comparing to Stockhardt et al. [15], who used three groups of German school students and provided one type of training per group, the false-negative rates decreased with a greater difference between the pre- and post-training results. This more substantial decrease might also be explained by the procedural choice of administering the testing immediately after the training. The differences between these groups before the training were large in both studies, which supports using larger group sizes in future studies.

4.2. Participant confidence

Based on the self-assessment, we observed a significant increase in participant confidence in their ability to detect phishing emails. However, there is no significant difference between the two training combinations.

Comparing to the prior study by Stockhardt et al. [15], the pre- and post-training levels of confidence are similar. Stockhardt et al. measured a significant difference that favored instructor-based training compared to text-based training, while our study did not show a difference for the methods. One reason may be due to different demographics.

4.3. Training satisfaction

The analysis of the participants' training experience suggests that Group B had a higher average rating over the three questions and a higher rating for enjoyment and learning compared to Group A, who rated the protection question slightly higher. The largest difference was in the learning amount question. However, none of these differences were statistically significant.

Comparing to Stockhardt et al. [15], their study also did not measure a significant difference in enjoyment. Thus, the results from both studies are similar as participants mostly enjoyed the training experiences.

The rating for "I think I learned a lot" also showed higher differences in the prior study as well as a similar result regarding the highest-rated instructor-based training and our Group B rating. Computer-based training and text-based training results were similar to the range of our Group A's result. One difference with the prior study is that the gap between the instructor-based training and the other training methods was significant.

When asked about the helpfulness of the training for better protecting themselves, we also observed results from our participants comparable to Stockhardt et al. Also, the averages of Group A in our study and the computer-based and text-based training in the prior study are similar as well as the averages of Group B in this study and instructor-based training in the prior study.

This comparison suggests that even though the demographics of the two studies were different, i.e., one was conducted with German students and the other with university students in Thailand, and the training was delivered through a single method in the prior study and a combination of methods in this study, the results are similar. Furthermore, based on the training experience question, while a slight favor appeared for instructor-led training, the difference was not significant in our study.

4.4. Training method preference

A strong preference for a specific training method for our groups of participants was not detected if being asked to rate each method regarding its helpfulness to educate them to be less susceptible to phishing attacks. Group A rated the video-based training highest over the text-based and game-based training. Group B also rated the text-based training similar to Group A, and game-based training slightly higher than text-based training, but lower than video-based training, while the classroom-based training was rated highest. This result indicates some favor for video-based training and more for classroom-based training. However, none of these ratings for each method differed significantly.

The final comparison asked participants to explicitly choose the single most-liked choice. Group A did not have a significant preference for any one method but rated video-based training the best, slightly ahead of game-based training. In Group B, the difference in favorability for classroom-based training was statistically significant when participants were asked to select only the single most-liked method.

Comparing to the prior study by Abawajy [20], we do see a difference in results. The prior study received the lowest rating for game-based training, more favor toward text-based training, and the highest favorability rating for video-based training. This contradicts our result from Group A which received a similar combination of training methods. A reason may be that the text-based training was different and that it was not in the native language of the participants in our study. The prior study did not include sufficient demographic information; thus, further comparison is not possible.

5. Conclusions

Based on these results, we conclude that the conducted security awareness training is generally effective in decreasing the false-negative rate while not increasing the false-positive rate. We also demonstrated that the training increases the participants' self-confidence. However, we also suggest that no specific training combination is significantly more effective than the other. In evaluating which training method combination achieved greater improvement, we determined that the outcome from the simulated phishing emails contradicted that of the screenshot assessment. We can conclude neither that additional classroom training provided a benefit in decreasing phishing susceptibility nor that it increased the participants' confidence in their ability to detect phishing. We showed that the participants have different opinions on the training methods with a majority preferring classroom-based training. However, the differences are not significant in all cases except if the participants are asked to choose one method only.

The contribution of this study provides confirmation that information security awareness programs should include a variety of learning methods to meet an audiences' range of preferences. For classroom-based training, our study does not provide justification from the perspective of improving effectiveness. Therefore, administrators of information

security awareness programs must balance the participants' preferences for classroom-based training with its more considerable effort required compared to other, less interactive methods. As preferences for training methods were heterogeneous in this and prior studies, we recommend first assessing the preferences for a target group to tailor the training program accordingly.

Limitations and further research

The results produced by the simulated phishing emails did not meet the requirements for the intended statistical methods, so their use for analysis is limited. One reason is that most participants rarely clicked on the simulated phishing emails and responded mostly to the phishing based on the "risk and loss" category. While these emails incorporated the look and feel of services provided by Mahidol University, an exact copy was not used nor was the university's name or logo, and the URLs were very different. The designs were closer to spear-phishing approaches by imitating a service known and utilized by all participants. Thus, a lesson learned and opportunity for future work is to investigate the usefulness of this measure further and compare it to the screenshot assessments by using a larger group of participants and sending many more emails with more effective designs (i.e., designs that closely mimic the services typically used by the participants). Future work should increase the size of participant groups to increase the opportunity for significance in the statistical tests based on the effect sizes of our power analyses in several tests. Additionally, future work could repeat the same approach with groups of alternate demographics. For example, it would be interesting to host a participant group with no IT background and greater age to compare the results of phishing susceptibility and training method preferences.

A limitation might be that we informed participants that they were participating in a study on phishing. This requirement may have resulted in participants being sensitized toward phishing emails and may have affected the results during the pre-training simulated phishing distribution. Although we did not have to inform them specifically about the simulated phishing emails and we were not approached by participants with questions or comments about these emails during the pre-training and training sessions, one participant contacted us during the post-training phase and asked if one of the four messages received by the person was from the research team. We informed the participant that it was sent from us and requested they not share this awareness with the others until after the study was complete. We have no reason to believe they shared the knowledge and have no indication that other participants became aware. However, future work may question participants afterward if they notice the emails and if they assumed they are part of the study.

As our study focuses on comparing combinations of methods used for training Thai participants and evaluating the possibility of measuring phishing susceptibility with simulated phishing emails, we reveal that phishing awareness training is beneficial for the participants, but classroom-based training does not necessarily increase effectiveness. We also identified that results from prior studies are mostly transferable to our demographic group and that further research on measurements using simulated phishing emails, as well as other demographic groups in Thailand, could be beneficial to improve security awareness training.

Declarations

Author contribution statement

Kai Florian Tschakert: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Sudsanguan Ngamsuriyaroj: Conceived and designed the experiments; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Funding statement

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Competing interest statement

The authors declare no conflict of interest.

Additional information

The screenshots of all stimuli used in the questionnaire as well as of the simulated phishing emails can be downloaded at https://awaresspaper.floriantschakert.de/.

Supplementary content related to this article has been published online at https://doi.org/10.1016/j.heliyon.2019.e02010.

References

[1] K. Rekouche, "Early Phishing", 4, CoRR, 2011 abs/1106.
[2] T. Hall, B. Hau, M. Penrose, M. Bevilacqua, Mandiant M-Trends 2016 EMEA Edition, 2016.
[3] Kaspersky Lab, "The Dangers of Phishing : Help Employees Avoid the Lure of Cybercrime,", 2015.
[4] Verizon, "2016 Data Breach Investigations Report,", 2016.
[5] Cyberark Software, "Global Advanced Threat Landscape Survey,", 2016.
[6] Europol, "Internet Organised Crime Threat Assessment 2016,", 2016.
[7] R. Dhamija, J.D. Tygar, M. Hearst, Why phishing works, in: In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems – CHI '06, 2006, p. 581. November 2005.
[8] M. Alsharnouby, F. Alaca, S. Chiasson, Why phishing still works: user strategies for combating phishing attacks, Int. J. Hum. Comput. Stud. 82 (2015) 69–82.
[9] P. Kumaraguru, S. Sheng, A. Acquisti, L.F. Cranor, J. Hong, Lessons from a real world evaluation of anti-phishing training, eCrime Res. Summit, eCrime 2008 (2008).
[10] C.B. Mayhorn, P.G. Nyeste, Training users to counteract phishing, Work 41 (SUPPL.1) (2012) 3549–3552.
[11] S. Sheng, M. Holbrook, P. Kumaraguru, L.F. Cranor, J. Downs, Who falls for phish? A demographic analysis of phishing susceptibility and effectiveness of interventions, in: Proc. 28th Int. Conf. Hum. Factors Comput. Syst. – CHI '10, 2010, pp. 373–382.
[12] A. Alnajim, M. Munro, An anti-phishing approach that uses training intervention for phishing websites detection, in: In ITNG 2009 – 6th International Conference on Information Technology: New Generations, 2009, pp. 405–410.
[13] S. Sheng, B. Magnien, Anti-phishing phil: the design and evaluation of a game that teaches people not to fall for phish, Proc. SOUPS (2007) 88–99, 2007.
[14] C.I. Canfield, B. Fischhoff, A. Davis, Quantifying phishing susceptibility for detection and behavior decisions, Hum. Factors 58 (8) (2016) 1158–1172.
[15] S. Stockhardt, B. Reinheimer, M. Volkamer, P. Mayer, A. Kunz, P. Rack, D. Lehmann, Teaching phishing-security: which way is best? in: J.-H. Hoepman, S. Katzenbeisser (Eds.), In ICT Systems Security and Privacy Protection: 31st IFIP TC 11 International Conference, SEC 2016, Ghent, Belgium, May 30 - June 1, 2016, Proceedings Springer International Publishing, Cham, 2016, pp. 135–149.
[16] J. Abawajy, T. Kim, Performance analysis of cyber security awareness delivery methods, Secur. Technol. disaster Recover. Bus. 122 (2010) 142–148.
[17] T. Caldwell, Making security awareness training work, Comput. Fraud Secur. 2016 (6) (2016) 8–14.
[18] Safety in Canada, What is Phishing? [Online]. Available: https://www.youtube.com/watch?v=9TRR6lHviQc, 2014. (Accessed 25 April 2017).
[19] AARP Academy, "What Is Phishing and How Do I Protect Myself,", 2016 [Online]. Available: https://www.youtube.com/watch?v=WpaLmeHTp3I. (Accessed 25 April 2017).
[20] J. Abawajy, User preference of cyber security awareness delivery methods, Behav. Inf. Technol. 33 (2014) 236–247. March 2016.
[21] G. Canova, M. Volkamer, C. Bergmann, R. Borza, "NoPhish: an anti-phishing education app,", Secur. Trust Manag. 10th Int. Work 8743 (2014) 88–192. STM 2014, Wroclaw, Poland, Sept. 10-11, 2014. Proc.
[22] G. Canova, M. Volkamer, C. Bergmann, R. Borza, B. Reinheimer, S. Stockhardt, R. Tenberg, Learn to spot phishing URLs with the android NoPhish app, in: M. Bishop, N. Miloslavskaya, M. Theocharidou (Eds.), In Information Security Education across the Curriculum: 9th IFIP WG 11.8 World Conference, WISE 9, Hamburg, Germany, May 26-28, 2015, Proceedings, Springer International Publishing, Cham, 2015, pp. 87–100.
[23] onlymyemail.com, "Facebook Phishing Catches Many," onlymyemail.com. [Online]. Available: http://blog.onlymyemail.com/facebook-phishing-catches-many/. (Accessed 11 May 2017).
[24] L. Muthiyah, "WHAT IS PHISHING? HOW TO CREATE PHISHING PAGE | FACEBOOK EXAMPLE," 7xter.Com, 2016 [Online]. Available: https://www.7xter.com/2016/08/phishing.html. (Accessed 11 May 2017).

[25] K. Parsons, A. McCormac, M. Pattinson, M. Butavicius, C. Jerram, The design of phishing studies: challenges for researchers, Comput. Secur. 52 (2015) 194–206.
[26] The DigiTrust Group, 10 Red Flags of Email Phishing, 2017 [Online]. Available: https://www.digitrustgroup.com/10-red-flags-email-phishing/. (Accessed 11 May 2017).
[27] ITS Information Security Office University of Otago, "Examples of Phishing Emails." [Online]. Available: https://blogs.otago.ac.nz/infosec/examples-of-phishing-emails/. (Accessed 11 May 2017).
[28] "K Kiwihead, Cyber Banking >>Phishing???, 2015 [Online]. Available: https://pantip.com/topic/34174350. (Accessed 11 May 2017).
[29] The Board of Governors of the Federal Reserve System, "You Can Fight Identity Theft.".
[30] SonicWall, "SonicWall Phishing IQ Test." [Online]. Available: https://www.sonicwall.com/phishing/.(Accessed 14 May 2017).
[31] OpenDNS, "PHISHING QUIZ." [Online]. Available: https://www.opendns.com/phishing-quiz/. (Accessed 14 May 2017).
[32] P. Finn, M. Jakobsson, Designing and conducting phishing experiments, IEEE Technol. Soc. Mag. Spec. Issue Usability Secur. (2007) 1–21.
[33] V. Anandpara, A. Dingman, M. Jakobsson, D. Liu, H. Roinestad, Phishing IQ tests measure fear, not ability, Financ. Cryptogr. Data Secur. (2007) 362–366.
[34] P. Kumaraguru, Y. Rhee, S. Sheng, S. Hasan, A. Acquisti, L.F. Cranor, J. Hong, Getting users to pay attention to anti-phishing education: evaluation of retention and transfer, APWG eCrime Res. Summit (2007) 70–81.
[35] P. Kumaraguru, S. Sheng, A. Acquisti, L.F. Cranor, J. Hong, Teaching Johnny not to fall for phish, ACM Trans. Internet Technol. 10 (2) (2010) 1–31.
[36] P. Kumaraguru, J. Cranshaw, A. Acquisti, L. Cranor, J. Hong, M.A. Blair, T. Pham, School of phish, in: Proc. 5th Symp. Usable Priv. Secur. - SOUPS '09, 2009, p. 1.
[37] D.M. Sarno, J.E. Lewis, C.J. Bohil, M.K. Shoss, M.B. Neider, Who are phishers luring?: a demographic analysis of those susceptible to fake emails, Proc. Hum. Factors Ergon. Soc. 2017 (2017) 1735–1739. Octob.
[38] J.G. Mohebzada, A. El Zarka, A.H. Bhojani, A. Darwish, "Phishing in a university community," 2012, Int. Conf. Innov. Inf. Technol. (2012) 249–254.
[39] D.M. Green, J.A. Swets, Signal Detection Theory and Psychophysics, John Wiley & Sons Canada, Limited, New York, 1966.
[40] D. Heeger, Signal Detection Theory, 1997, pp. 1–10.
[41] S. Chartier, D. Cousineau, Computing mixed-design (Split-Plot) ANOVA, Math. J. 13 (2011) [Online]. Available: http://www.mathematica-journal.com/2011/10/computing-mixed-design-split-plot-anova/. (Accessed 21 April 2017).
[42] G. Aaron, R. Rasmussen, A. Routt, "Global Phishing Survey: Trends and Domain Name Use in 1H2014," Anti-phishing Work. Gr. (APWG), 1, Phishing Rep. H, Lexington, MA, 2014, p. 2014 [Online]. Available, "https://docs.%20apwg.%20org/reports/APWG%20Glob".
[43] R. Hall, 2x2 Mixed Factorial Design, 1998 [Online]. Available: https://web.mst.edu/~psyworld/mixed_designs.htm. (Accessed 21 April 2017).
[44] Laerd Statistics, "Two-Way Mixed ANOVA Using SPSS Statistics," Statistical Tutorials and Software Guides, 2015 [Online]. Available: https://statistics.laerd.com/premium/spss/twma/two-way-mixed-anova-in-spss.php. (Accessed 10 October 2017).
[45] Laerd Statistics, "McNemar's Test in SPSS Statistics," Statistical Tutorials and Software Guides, 2015 [Online]. Available: https://statistics.laerd.com/premium/spss/mt/mcnemars-test-in-spss.php. (Accessed 31 January 2019).
[46] Laerd Statistics, "Mann-Whitney U Test in SPSS Statistics," Statistical Tutorials and Software Guides, 2015 [Online]. Available: https://statistics.laerd.com/premium/spss/mwut/mann-whitney-test-in-spss.php. (Accessed 31 January 2019).
[47] L.C. Dinneen, B.C. Blakesley, [Algorithm AS 62] A generator for the sampling distribution of the Mann-Whitney U statistic, Appl. Stat. 22 (1973) 269–273.
[48] Laerd Statistics, "Kruskal-Wallis H Test in SPSS Statistics," Statistical Tutorials and Software Guides, 2015 [Online]. Available: https://statistics.laerd.com/premium/spss/kwht/kruskal-wallis-test-in-spss.php. (Accessed 31 January 2019).
[49] Laerd Statistics, "Chi-Square Goodness-Of-Fit Using SPSS Statistics," Statistical Tutorials and Software Guides, 2015 [Online]. Available: https://statistics.laerd.com/premium/spss/gof/goodness-of-fit-in-spss.php. (Accessed 18 October 2017).
[50] R-core Team, R-core@R-project.org, "Chisq from the R Stats Package v3.5.2,", 2015, p. 31 [Online]. Available: https://www.rdocumentation.org/packages/stats/versions/3.5.2/topics/chisq.test. (Accessed January 2019).
[51] The R Foundation, "R: What is R?" [Online]. Available: https://www.r-project.org/about.html. (Accessed 24 May 2019).
[52] S.S. Mangiafico, "R Companion: Exact Test of Goodness-Of-Fit,", 2015 [Online]. Available: http://rcompanion.org/rcompanion/b_01.html. (Accessed 31 January 2019).