future internet

MDPI

Article

Evaluation of Contextual and Game-Based Training for Phishing Detection

Joakim Kävrestad [1,*], Allex Hagberg [2], Marcus Nohlberg [1], Jana Rambusch [1], Robert Roos [2] and Steven Furnell [3]

[1] School of Informatics, University of Skövde, 541 28 Skövde, Sweden; marcus.nohlberg@his.se (M.N.); jana.rambusch@his.se (J.R.)
[2] Xenolith AB, 541 34 Skövde, Sweden; allex@xenolith.se (A.H.); robert@xenolith.se (R.R.)
[3] School of Computer Science, University of Nottingham, Nottingham NG7 2RD, UK; steven.furnell@nottingham.ac.uk
* Correspondence: joakim.kavrestad@his.se

Abstract: Cybersecurity is a pressing matter, and a lot of the responsibility for cybersecurity is put on the individual user. The individual user is expected to engage in secure behavior by selecting good passwords, identifying malicious emails, and more. Typical support for users comes from Information Security Awareness Training (ISAT), which makes the effectiveness of ISAT a key cybersecurity issue. This paper presents an evaluation of how two promising methods for ISAT support users in acheiving secure behavior using a simulated experiment with 41 participants. The methods were game-based training, where users learn by playing a game, and Context-Based Micro-Training (CBMT), where users are presented with short information in a situation where the information is of direct relevance. Participants were asked to identify phishing emails while their behavior was monitored using eye-tracking technique. The research shows that both training methods can support users towards secure behavior and that CBMT does so to a higher degree than game-based training. The research further shows that most participants were susceptible to phishing, even after training, which suggests that training alone is insufficient to make users behave securely. Consequently, future research ideas, where training is combined with other support systems, are proposed.

Keywords: usable security; cybersecurity training; ISAT; SETA; phishing; user awareness; security behavior

Citation: Kävrestad, J.; Hagberg, A.; Nohlberg, M.; Rambusch, J.; Roos, R.; Furnell, S. Evaluation of Contextual and Game-Based Training for Phishing Detection. *Future Internet* **2022**, *14*, 104. https://doi.org/10.3390/fi14040104

Academic Editors: Leandros Maglaras, Helge Janicke and Mohamed Amine Ferrag

Received: 7 March 2022
Accepted: 22 March 2022
Published: 25 March 2022

1. Introduction

The world is continuing a journey towards an increasingly digital state [1]. The use of computers and online services has been a natural component of the lives of most people in developed countries for decades and adoption in developing regions is on the rise [2]. Furthermore, populations that previously demonstrated low adoption rates are now adopting and using digital services at a rapid pace [3,4]. This development is positive. On a national level, Internet adoption has been shown to positively impact financial development [2]. On the individual level, the use of digital services makes it easier for the individual to access information, healthcare, and more, while enabling social contact in situations where meeting physically is challenging or even impossible [5,6].

However, digitalization is not without risk. The move to more digital work, leisure and more also means a move to more digital crime and threats [7]. Digital threats expose users and organizations to risks daily, and the need for cybersecurity to protect against those risks is undeniable. The threat landscape is multi-faceted and includes various types of threats that can be broadly classified as technological or human [8]. Technological threats include, for instance, malware or hacking where the attacker is using technological means to destroy or gain access to devices or services. Human threats involve exploiting user behavior, typically for the same purpose. A common type of human threat is phishing,

where an attacker sends an email to the target victim and attempts to persuade the victim into behaving in an insecure way by, for instance, downloading an attachment or clicking a link and then submitting login credentials to some service. Phishing is continuously reported as the most common threat to both organizations and individuals, and therefore the topic of this paper [9–11].

At its core, phishing is when an attacker attempts to trick a user into insecure behavior. Insecure behavior typically includes downloading a malicious attachment, clicking a link or giving up sensitive information in reply to the email [12]. Phishing has traditionally been easy to spot as generic messages which are often poorly formatted with poor spelling and grammar [13]. While that is still true for some of today's phishing campaigns, now many phishing emails are well-written and use various techniques to invoke trust [12]. Furthermore, attackers employ targeted attacks where they tailor emails to a specific recipient, a technique known as spear-phishing [9]. In such an attack, the attacker may steal the email address of a friend or coworker of the target victim and make the email appear to come from that known sender. The attacker may also research the victim and ensure that the content of the malicious email is content that the victim would, given the victim's job position or interest, expect to receive [14].

Techniques used by attackers and techniques used to defend against phishing both include technical and human aspects [15]. An attacker will exploit human behavior to invoke trust and persuade the victim into insecure behavior. As part of the attack, the attacker may also exploit technical weaknesses in the email protocols to pose as a trusted sender or use another technical weakness to take control of the victim's system once the victim opens a malicious attachment [12]. Likewise, several organizations employ technical measures, such as automatic filters, to defend against phishing. However, educating users on detecting phishing emails remains the most commonly suggested defense mechanism. While both technical and human aspects of phishing are important, the primary focus of this paper is on the human side, particularly on user behavior and how it can be understood and improved.

As explained by the knowledge, attitude, and behavior (KAB) model, behavior is influenced by knowledge, and attitude [16]. KAB describes that increased knowledge about an expected behavior will lead to increased awareness and, finally, a change in behavior. This relationship has been evaluated in the security domain and found to hold [17].

Information Security Awareness Training (ISAT) is commonly suggested as the way to improve user awareness [18–20]. There are several different ways to train users presented in the literature. These include providing lectures, text-based warnings, video instructions sent out via email at regular intervals, instructive games and training automatically provided to users in risky situations [21–25]. There are, however, several publications suggesting that many training efforts fail to support users towards secure behavior to a high enough degree [26,27]. Suggested reasons include that it is hard to make users participate in on-demand training, that acquired knowledge is not retained for long enough, and that knowledge does not necessarily translate to correct behavior [20,28]. Some research even suggests that training methods are not empirically evaluated to a high enough extent [29,30].

This paper seeks to evaluate the effectiveness of two promising methods for ISAT; game-based training and Context-Based Micro-Training (CBMT). Game-based training means that users are presented with an educative game and is argued to increase user participation rates and provide a more realistic training environment compared to lectures, videos, or similar [31]. CBMT means that users are presented with condensed information in situations where the training is of direct relevance. In the context of phishing, a user will receive training when opening a mailbox. CBMT is argued to increase users' awareness and has been evaluated in the context of password security with positive results [32]. The research question addressed in this paper is:

> To what extent can the two methods, game-based training and CBMT, support users to accurately differentiate between phishing and legitimate email?

The research was carried out as a simulated experiment with 41 participants. The participants were asked to identify phishing emails while their behavior was monitored using an eye-tracking technique. The results show that both training methods can support users towards secure behavior and that CBMT does so to a higher degree than game-based training, which makes the first contribution of this research. The research further shows that most participants were susceptible to phishing, even after training which suggests that training alone is not enough to make users behave securely. The upcoming section will elaborate on ISAT and justify the selection of CBMT and game-based training as the focus of this research. The rest of this paper will, in turn, present the research methodology results, and discuss those results and their limitations.

2. Information Security Awareness Training

ISAT has been discussed in the scientific literature for several decades, and the importance of providing ISAT as a means of improving user behavior is widely acknowledged [33–35]. ISAT intends to increase user knowledge and awareness through training. There are many and diverse, options for ISAT, and recent publications [35–37] categorize ISAT methods differently. In general terms, ISAT methods can be described as seen in Table 1. Table 1 is based on the classifications by [36,37].

Table 1. Overview of ISAT methods.

Method	Description
Classroom training	Typically provided on-site as a lecture attended as a specific point in time.
Broadcasted online training	Typically, brief training delivered as broadcast to large user groups using e-mail or social networks.
E-learning	ISAT typically delivered using an online platform that is accessible to users on-demand.
Simulated or contextual training	Training delivered to users during a real or simulated event.
Gamified training	Gamified training is described as using gamification to develop ISAT material.

While ISAT has been long discussed in scientific literature and used in practice, several publications suggest that many ISAT methods fail to adequately support users towards secure behavior [26,27]. This notion is emphasized by the continuous reports of incidents where human behavior is a key component [38,39]. Three core reasons for why ISAT does not always provide its intended effect can be found in recent research:

- Knowledge acquired during training deteriorates over time [21].
- It is challenging to get users to participate in training delivered on-demand [28].
- Users are provided with knowledge, but not motivated to act in accordance to that knowledge [20].

The ISAT methods included in this research are game-based training and Context-Based Micro-Training (CBMT). Gamified training means that game concepts are applied to ISAT, with the intent to better motivate users to actively participate [28]. It is considered in this research since it is argued to better motivate and engage users when compared to other ISAT alternatives. There are several examples of gamified ISAT. The landscape includes multi-player competitive games, story-based single-player games, board games, role-playing games, quizzes, and more [28,40].

CBMT is an example of contextual training. ISAT using the CBMT method is delivered to users in short sequences and in situations where the training is of direct relevance. Phishing training is, for instance, delivered to users that are in a situation with an elevated risk of being exposed to phishing. It is argued to counter the knowledge retention and user

participation problems by automatically appearing in those relevant situations [32]. It is also argued to motivate users towards secure behavior by providing them with training that directly relates to the users' current situation.

3. Materials and Methods

The purpose of this study was to evaluate user behavior when assessing if emails are malicious or not. To that end, a controlled experiment where the participants were exposed to an inbox and asked to classify the email contained in that inbox was conducted. The participants were scored based on how accurately they classified the emails. Furthermore, the participants' behavior was monitored during the experiment by an eye tracker that recorded where the participants were looking on screen. Before the experiments, the participants were randomised into three groups; game-based training, CBMT-based training or control. A between-group analysis was performed to identify differences between training methods and answer the research question posed. As detailed at the end of paper statements, data supporting this paper is available as open data (https://doi.org/10.5878/g6d9-7210 (accessed on 6 March 2022)). Furthermore, the study did not require ethical review, but all participants signed a written informed consent form detailing how the study was executed and how data were handled. An overview of the research process is presented in Figure 1. The rest of this section provides a detailed description of the experiment environment, data collection procedures, collected variables, and data analysis procedures.

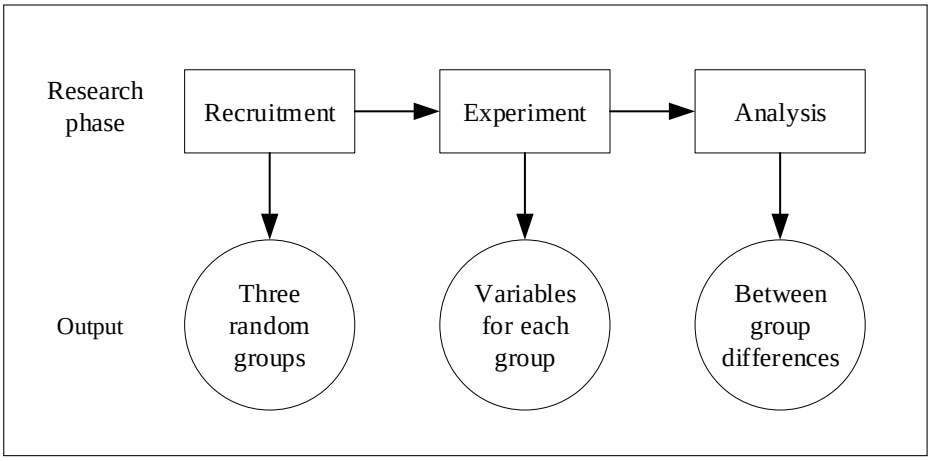

Figure 1. Research process overview.

3.1. Experiment Environment

An experiment environment containing an email system was set up on Ubuntu Linux using the email server and management platform Modoboa (https://modoboa.org/en/ (accessed on 6 March 2022)). Both Ubuntu Linux and Modoboa were installed with default settings. Modoboa allowed for the creation of unlimited email domains and addresses and provided a webmail interface. Several email domains were configured so that different types of emails could be created:

- Legitimate emails from service providers such as Google and banks.
- Phishing emails that imitated phishing emails from hijacked sender accounts.
- Phishing emails from domains made up to look similar to real domains, for instance, lundstro.mse instead of lundstrom.se.

The fictitious company Lundström AB, and the character Jenny Andersson were developed. The company was given the domain lundstrom.se and the character was given the email address jenny@lundstrom.se. A persona was developed for Jenny Andersson. The experiment participants were asked to assume Jenny´s persona and classify the email in her inbox. The persona was expressed as follows:

> Jenny is 34 years old and works as an accountant at a small company (Lundström AB), and her manager is Arne Lundtröm. She lives with her husband and kids in a small town in Sweden. Your email address is jenny@lundstrom.se. You use the banks SBAB and Swedbank and is interested in investing in Bitcoin. You are about to remodel your home and have applied for loans at several banks to finance that. You shop a fair bit online and are registered at several e-stores without really remembering where. You are currently about to remodel your bathroom. Ask the experiment supervisor if you need additional information about Jenny or the workplace during the experiment.

Jenny's inbox was populated with 11 emails where five were legitimate, and six were phishing. The legitimate emails were crafted as reasonable questions from her manager or communications from banks and craftsmen. The communications from banks and craftsmen were based on real emails taken from one of the researcher's inboxes. The six phishing emails were crafted to include different phishing identifiers. Five different phishing identifiers were included in the experiment. They are commonly mentioned in scientific and popular literature and were the following [41–44]:

1. Incorrect sender address where the attacker may use an arbitrary incorrect sender address, attempt to create an address that resembles that of the true sender, or use a sender name to hide the actual sender address.
2. Malicious attachments where the attacker will attempt to make the recipient download an attachment with malicious content. A modified file extension may disguise the attachment.
3. Malicious links that are commonly disguised so that the user needs to hover over them to see the true link target.
4. Persuasive tone where an attacker attempts to pressure the victim to act rapidly.
5. Poor spelling and grammar that may indicate that a text is machine translated or not written professionally.

The included phishing emails are described as follows:

1. The first phishing email came from the manager's real address and mimicked a spear-phishing attempt, including a malicious attachment and hijacked sender address. The attachment was a zip file with the filename "annons.jpeg.zip (English: advertisement.jpeg.zip)". The text body prompted the recipient to open the attached file. In addition to a suspicious file extension, the mail signatures differed from the signature in other emails sent by the manager.
2. The second phishing email came from Jenny's own address and prompted the recipient to click a link that supposedly led to information about Bitcoin. The email could be identified as phishing by the strange addressing and the fact that the tone in the email was very persuasive.
3. The third phishing email appeared to be a request from the bank SBAB. It prompted the user to reply with her bank account number and deposit a small sum of money into another account before a loan request could be processed. It could be identified by improper grammar, an incorrect sender address (that was masked behind a reasonable sender name), and the request itself.
4. The fourth phishing email was designed to appear from Jenny´s manager. It prompted Jenny to quickly deposit a large sum of money into a bank account. It could be identified by the request itself and because the sender address was arne@lundstro.mse instead of arne@lundstrom.se.
5. The fifth phishing email mimicked a request from Google Drive. It could be identified by examining the target of the included links that lead to the address xcom.se instead of google.
6. The sixth phishing email appeared to be from the bank Swedbank and requested the recipient to go to a web page and log in to prove the ownership of an account. It could

be identified as phishing by examining the link target, the sender address, which was hidden behind a sender name, and the fact that it contained several spelling errors.

The experiment was set up so that most phishing emails had similar legitimate counterparts. The legitimate emails included where:

1. The first legitimate email was a request from Jenny's manager Arne. The request prompted Jenny to review a file on a shared folder.
2. The second legitimate email was a notification from a Swedish bank. It prompted Jenny to go to the bank website and log in. It did not contain any link.
3. The third legitimate email was an offering from a plumber. While containing some spelling errors, it did not prompt Jenny to make any potentially harmful actions.
4. The fourth legitimate email is a request for a meeting from Jenny's manager Arne.
5. The fifth email is a notification from a Swedish bank. This notification prompts the user to go to the bank website and log in. It does not contain any greeting or signature with address.

The webmail interface is demonstrated in Figure 2. Figure 2 displays the layout of the included emails and is annotated to show the ordering of the emails. Legitimate emails are denoted L*n*, in green, and phishing emails are denoted P*n* in red.

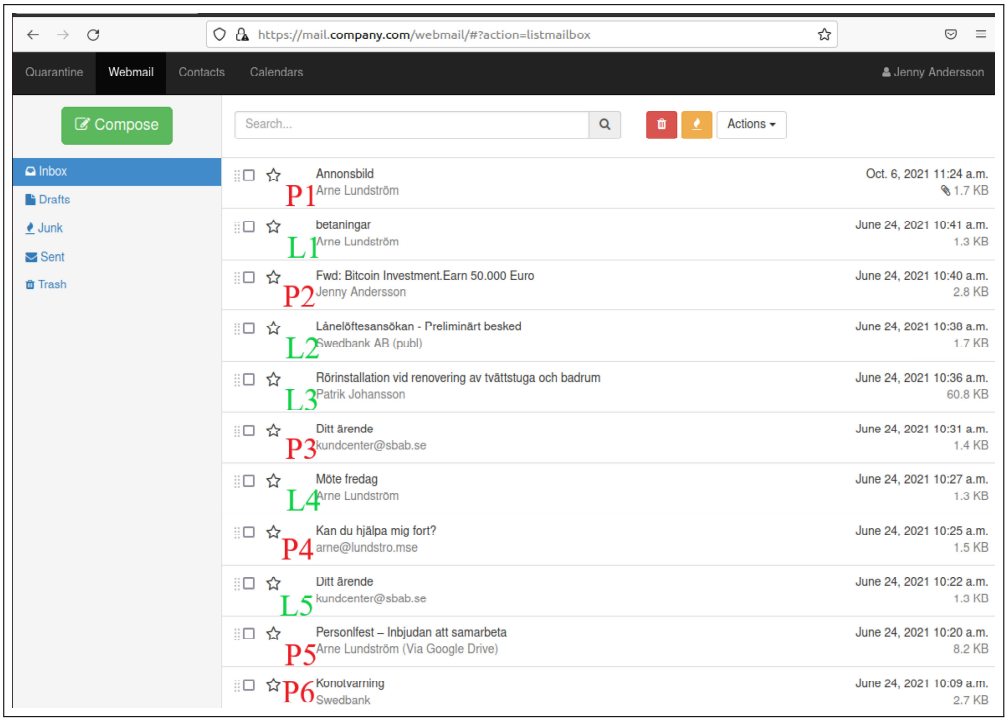

Figure 2. Webmail interface used in the experiment.

3.2. Participant Recruitment

Participants were recruited using a convenience sampling approach where students and employees from the University of Skövde were recruited. Participants with education or work experience in cybersecurity were excluded from the study. All participants were invited with a direct email that they were asked to reply to in order to participate. Upon registration, participants were randomly assigned to one of the three groups and provided with a description of the experiment, a description of the persona, and an informed consent form. The three groups were the following:

- Game: Participants in this group were prompted to play an educational game before arriving for the experiment. The game is called Jigsaw (https://phishingquiz.withgoogle.com/) (accessed on 6 March 2022) and is developed by Google. It is an example of game-based training that is implemented as a quiz and was selected for use

in this research because it is readily available for users. It also covers all the identifiers of phishing previously described. Jigsaw takes about five minutes to complete.

- CBMT: Participants in this group received computerized training developed by the research team according to the specifications of CBMT. It was written information that appeared to the participants when they opened Jenny's inbox, as demonstrated in Figure 3. The participants were presented with a few tips and prompted to participate in further training, which led the participants to a text-based slide show in a separate window. The training takes about five minutes to complete.
- CONTROL: This group completed the experiment without any intervention.

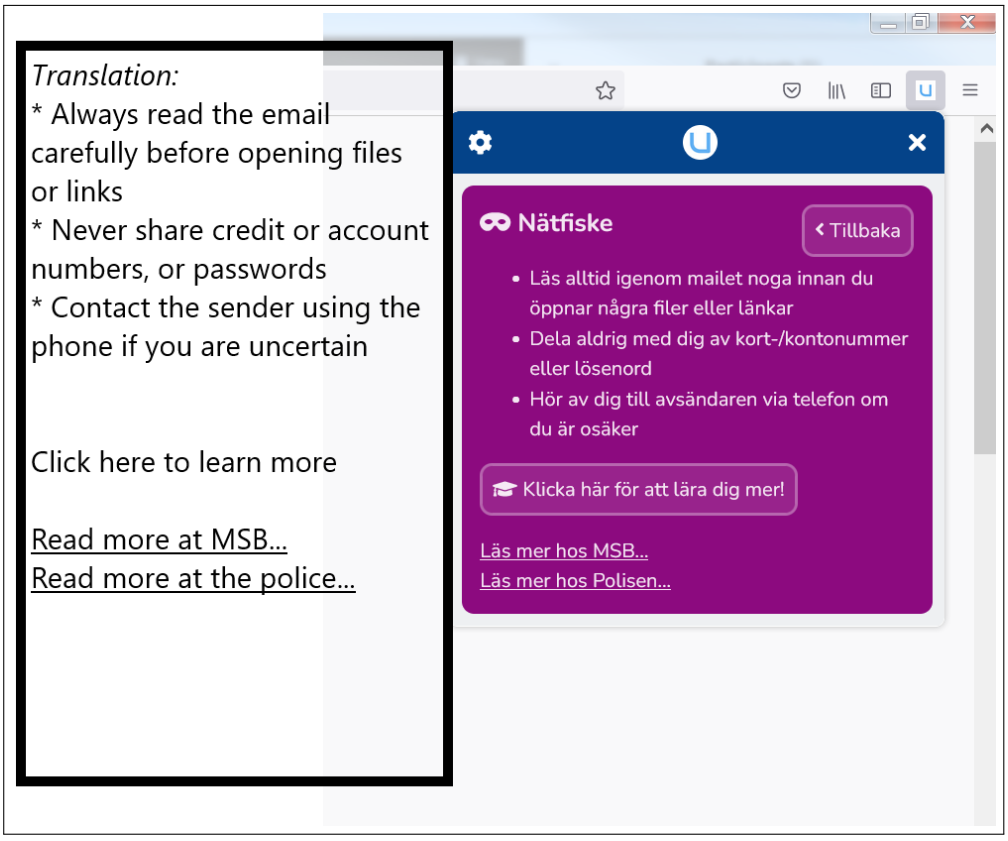

Figure 3. Demonstration of CBMT-based training.

3.3. Experiment Procedure

On arriving for the experiment, the participant was seated in a regular office in front of a 24″ computer monitor that displayed the experiment environment. The monitor was equipped with a Gazepoint GP3 HD eye tracker (https://www.gazept.com/product/gp3hd/) (accessed on 6 March 2022). The participant was asked to read the informed consent form and given the opportunity to ask questions about the experiment and study before signing it. The participant was then asked to respond to a survey with demographic questions and asked to take a seat in front of the monitor. The eye tracker was calibrated using the manufacturer's built-in calibration sequence with nine points [45]. The calibration was considered successful when the control software deemed all nine points valid. In cases where the eye tracker could not be successfully calibrated, eye-tracking data were disregarded for that participant. This happened for three participants.

The participant was then reminded of Jenny's persona and asked to classify the email in Jenny's inbox. The participant was instructed to delete all phishing emails and keep all legitimate emails. The participant was asked to think aloud during the experiment, especially about how decisions to delete emails were made. The participant was also told that at least one of the emails was phishing and that a score was to be calculated based on the participants' performance. The intent was to make the participant as aware

of phishing as possible. The rationale was that mere inclusion in the experiment would increase the participant's awareness level, and by priming all participants to high awareness would make the awareness levels of the participants comparable. Consequently, the gathered data reflects the participants' best ability to delete phishing rather than the ability they can be assumed to have during their daily work. Gazepoint analysis UX Edition (https://www.gazept.com/product/gazepoint-analysis-ux-edition-software/) (accessed on 6 March 2022) was used to monitor the participant's performance in real time on an adjacent screen and for post-experiment analysis of the collected eye-tracking data. Following the experiment, the participants in the game group were asked if they had played the game before the experiment as instructed. The experiment process, from the participant's point of view, is visualized in Figure 4.

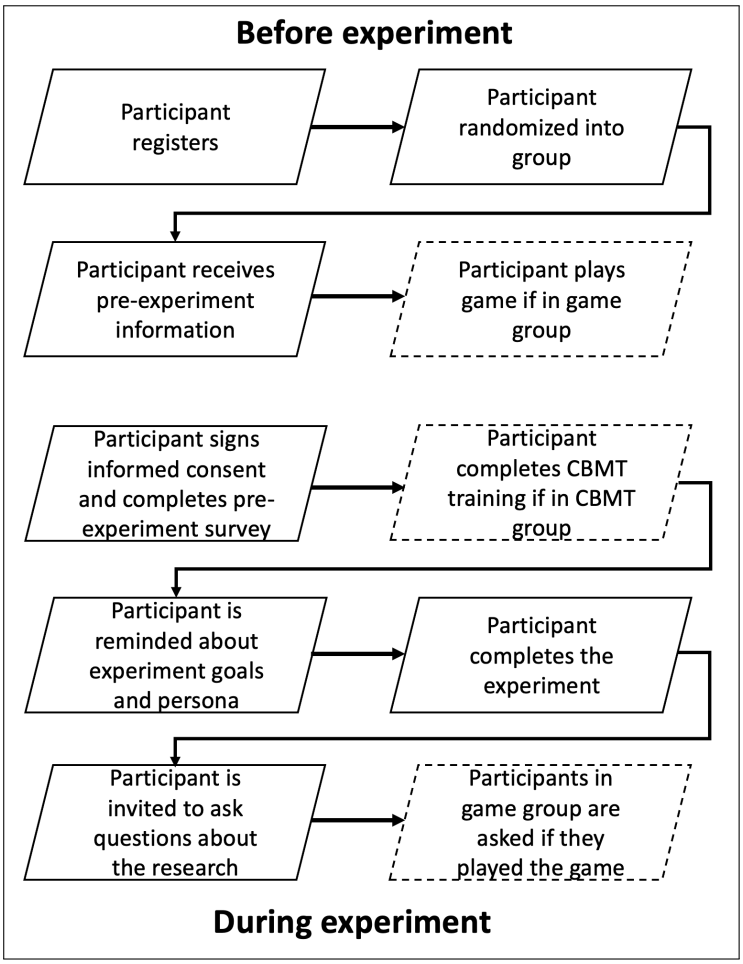

Figure 4. Visualization of experiment procedure. Dashed boxes only applied to some groups.

3.4. Collected Variables

Variables reflecting the participants' demographic background, score and behavior were captured during the experiment. The demographic variables were collected to enable a descriptive presentation of the sample's demographic attributes. The score variables reflected the total number of correct classifications the participants made. The behavior variables described how the participants acted during the experiment by counting how many of the previously described phishing identifiers the participants used. Two behavior variables were collected. The first was collected manually during the experiment (behavior_manual). It was based on real-time monitoring, and the participants expressed thoughts. It reflected how many of the following actions the participant performed at least once:

1. Evaluated the sender address by hovering over the displayed name to see the real sender address.
2. Evaluated attachments by acknowledging their existence and describing it as suspicious or legitimate.
3. Evaluated links by hovering over them or in some other way verified the link destination.
4. Evaluated if the tone in the email was suspiciously persuasive.
5. Evaluated if spelling and grammar made the email suspicious.

Please note that the variables reflect what identifiers the participants used but not if they accurately interpreted the information provided by the identifier. A participant who, for instance, incorrectly evaluated a sender address as legitimate would still get the point for evaluating the sender address. The second behavior variable, behavior_tracked, was computed automatically by defining Areas of Interest in Gazepoint analysis UX Edition and counting how many times the participant gazed in those areas. Areas of Interest are defined screen areas that allow for collecting the number of times the participants gaze in those particular areas. The following three Areas of Interest were defined.

- Address, which covered the area holding the sender and recipient addresses.
- Attachment covering the area where email attachments are visible.
- Link covering the area where the true link destination appears.

The Areas of Interest were only active when they included the intended information. For instance, the Attachment area was only active when an attachment was visible on the screen. The Areas of Interest are demonstrated in Figure 5 which also shows how red dots denote where the participant is currently looking.

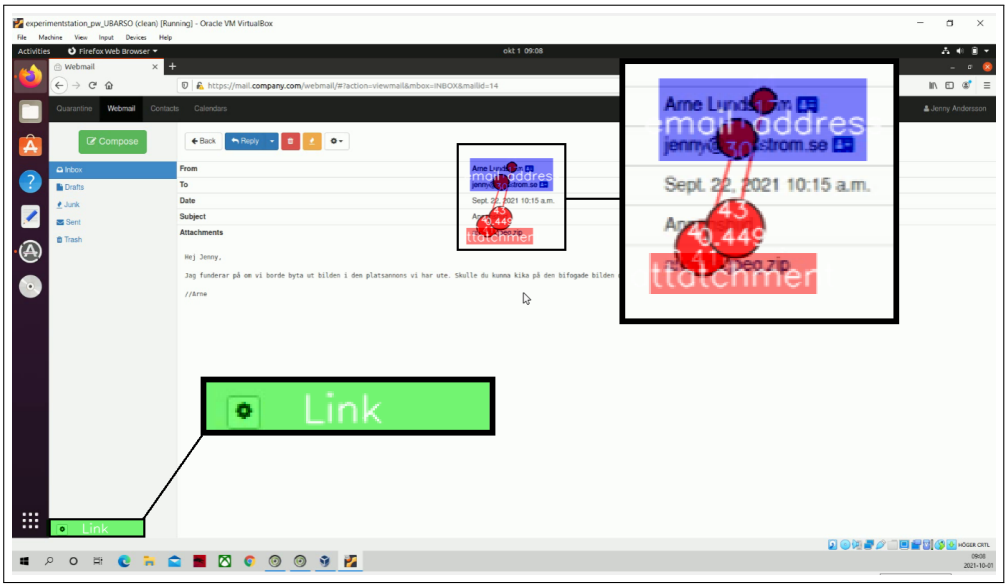

Figure 5. Demonstration of how areas of interest were defined, with AOI definitions enlarged. Please note that the Link area contains the target address of a link that is hovered over.

3.5. Data Analysis

The data were analyzed using SPSS version 25. The demographic properties of the sample were first described followed by a descriptive overview of the three variables SCORE, behavior_manual, and behavior_tracked. The proportion of participants that received perfect scores was then reported. A perfect score means that a participant identified all 11 emails correctly, or used all phishing identifiers assessed by the variables behavior_manual and behavior_tracked, respectively.

Next, Kruskal–Wallis H tests were used, with pairwise Mann–Whitney U test with Bonferroni correction as post hoc procedure, to identify significant between-group differ-

ences. Kruskal–Wallis H test performed on three or more samples will return a significant result if at least one sample is different from the others. In such a case, the Mann–Whitney U test with Bonferroni correction is used between all pairs in the sample to analyze what individual samples that are different from each other. Kruskal–Wallis H test was used over ANOVA because the data must show a normal distribution for ANOVA to be robust, and most samples did not in this case [46]. The conventional significance level of 0.05 is used throughout this paper.

4. Results

This section outlines the results of the study. It is divided into two sections were the first section outlines a descriptive overview of the data. The second section outlines the results in relation to the research question. It should be noted that three participants in the Game group reported that they did not play the provided game. This is to be expected given previous works suggesting that it is challenging to get users to participate in training [28]. All statistical procedures have been performed with and without those three participants. Results concerning the Game group are reported as *n(m)* were *n* is the result when the complete group is considered and *m* is the result when participants that did not play the game are omitted.

4.1. Data Overview

Data was collected over a period of about two months and included 41 participants. Two participants were removed from the data set since they reported having formal training in cybersecurity. The data collection period was intended to be longer, but data collections stopped after a security incident where the IT department warned all students and staff at the university about phishing involving attachments. Continued data collection would have risked the validity of the data set. The mean participant age was 37. Twenty-three participants identified themselves as female and 16 as male. Twenty-three participants reported being employees and 16 reported being students. An overview of the mean and median values for the collected variables and the distribution form of the variables is presented in Table 2. Please note that eye-tracking failed for three participants and the participants included for the variable behavior_tracked is therefore only 36.

Table 2. Data overview.

Variable	Group	Mean	Median	Normal Distribution
SCORE out of 11	Control (n = 11)	8.82	9	YES
	CBMT (n = 14)	10	10	NO
	Game (n = 14)	8.86 (9.09)	9 (9)	NO
	Total (n = 39)	9.26	9	NO
behavior_manual out of 5	Control (n = 11)	3	3	NO
	CBMT (n = 14)	4.57	5	NO
	Game (n = 14)	3.64 (3.82)	3.5 (4)	NO
	Total (n = 39)	3.79	4	NO
behavior_tracked out of 3	Control (n = 10)	1.9	2	NO
	CBMT (n = 12)	2.5	3	NO
	Game (n = 14)	2.29 (2.55)	2 (3)	NO
	Total (n = 36)	2.25	2	NO

4.2. The Effect of Training

The effect of training was assessed by first examining the proportion of participants that received perfect scores. A perfect score means that the participants used all phishing identifiers or identified all emails correctly. The proportions of perfect scores are presented in Table 3.

Table 3 suggests that participants who received training performed better than participants in the control group for the behavior variables and that the participants in the CBMT

group outperformed the other groups for the variable SCORE. The same tendency is seen in Table 2 where mean and median results for the different sample groups are presented. Table 2 suggests that participants in the group game performed slightly better than the control group while the participants in the group CBMT outperformed the other groups with a bigger margin. The exception is for the variable behavior_tracked where the groups CBMT and game performed equally when participants who reported not playing the game were omitted from the game group.

Table 3. Proportions of perfect scores.

Variable	Group	Perfect Scores
SCORE	Control (n = 11)	0%
	CBMT (n = 14)	21.4%
	Game (n = 14)	0% (0%)
	Total (n = 39)	7.7% (8.3%)
behavior_manual	Control (n = 11)	0%
	CBMT (n = 14)	64.3%
	Game (n = 14)	14.3% (18.2%)
	Total (n = 39)	28.2% (30.6%)
behavior_tracked	Control (n = 10)	9.1%
	CBMT (n = 12)	57.1%
	Game (n = 14)	42.9% (54.5%)
	Total (n = 36)	38.5% (45.5%)

Kruskal–Wallis H test was used to identify variables with statistically significant between-group differences. The results are presented in Table 4.

Table 4. Kruskal–Wallis H tests.

Variable	Kruskal–Wallis H	*p*-Value
SCORE	13.965 (12.531)	0.001 (0.002)
behavior_manual	16.270 (15.434)	0.000 (0.000)
behavior_tracked	5.569 (7.332)	0.062 (0.026)

The Kruskal–Wallis H tests suggest that at least one sample is different from the others when $p < 0.05$, as is the case for the variables SCORE and behavior_manual. The same is also true for the variable behavior_tracked when users who did not play the game are omitted. Pairwise Mann–Whitney U tests with Bonferroni correction was used to test what variables that were significantly different from each other. The results are presented in Table 5.

Table 5. Pairwise post hoc tests. Please note that post hoc tests for the variable behavior_tracked were only computed in the case when participants in the group Game, who did not play the game was omitted because the corresponding Kurskal-Wallis H tests was only significant in that case.

Variable	Groups	*p*-Value
SCORE	Control-Game	1.000 (1.000)
	Control-CBMT	0.005 (0.003)
	Game-CBMT	0.003 (0.023)
behavior_manual	Control-Game	0.502 (0.277)
	Control-CBMT	0.000 (0.000)
	Game-CBMT	0.021 (0.102)
behavior_tracked	Control-Game	X (0.083)
	Control-CBMT	X (0.036)
	Game-CBMT	X (1.000)

In this case, the difference between two variables is statistically significant when $p < 0.05$. Table 5 shows that CBMT is separated from the groups game and control for the variables SCORE and behavior_manual while control and game cannot be separated. For behavior_tracked, game and CBMT cannot be separated but are both separated from control.

5. Discussion

This research explores how effectively Information Security Awareness Training (ISAT) can support users to accurately identify phishing emails. The research evaluated two methods that were discussed as being promising in recent literature, namely game-based training and training based on CBMT. The research was conducted as a simulated experiment that measured how the participants behaved when assessing whether emails were phishing or not, and how accurately they classified email. The statistical analysis shows that participants in the CBMT group had higher scores than users in the game or control group. In terms of behavior, participants in the CBMT group performed better than the game and control group for the manually collected variable. However, the CBMT and game groups were equally strong for the variable computed based on eye-tracking data. In conclusion, both game-based training and CBMT are shown to improve user behavior in relation to phishing while only CBMT can be shown to improve users' ability to accurately classify phishing emails.

One reason could be that CBMT provides an awareness increasing mechanism in addition to training while game-based training does not. The game-based training is delivered to participants on a regular basis and was mimicked in the experiment by letting the participants take the training prior to arriving for the experiment. CBMT is, by design, presented to users when they are entering a risky situation and that was mimicked by presenting the CBMT training to participants just before starting the experiment. The difference in how the training was delivered could account for the difference in results between the two groups. In fact, the effect of awareness increasing mechanisms have been evaluated in prior research with good results [47,48]. This research extends those results by suggesting that awareness increasing mechanisms combined with training are likely to have a positive effect on users' ability to accurately identify phishing emails.

While training was proven to improve participants' ability to identify phishing, it can be noted that less than 10% of the participants were able to identify all emails correctly. Furthermore, less than 50% of the participants evaluated all of the phishing identifiers and even if the participants in the CBMT group received training just before starting the experiment, 35.7% of those participants missed one or more phishing identifiers. Yet, most organizations explicitly or implicitly expect users to correctly identify all phishing emails all the time. The present research shows that even if users are provided with training just before being tasked with identifying phishing, and instructed to actively search for phishing, very few users are able to fulfill the expectations of that security model. The implication of this result is that the security model or the feasibility of using training alone to reach it must be questioned. One could, for instance, question if we should follow a paradigm where users are expected to change according to how computers work. A more useful paradigm could be to modify the way that computers work to match the abilities of the users. A similar viewpoint is presented by [49] who questions why the responsibility for cybersecurity is individualized through the notion of the "stupid user". Instead, ref. [49] suggest that user-oriented threats should be managed by security professionals, and managers, at a collective level. Likewise, ref. [50] calls for a more holistic approach to anti-phishing methods.

5.1. Limitations

A given limitation of this study comes from participation bias. Participation bias is known to impact simulated experiments in cybersecurity awareness [22]. The expected effect in this study is that participants are more aware than they would be in a naturally occurring situation. Thus, the scores are expected to reflect the participants' best ability

rather than their average performance. Using a between-group design, we still argue that differences between ISAT methods identified in this research are valid. However, it is likely that the actual performance of the included methods will be lower in a natural environment. On a similar note, the method cannot account for organizational factors such as leadership support and social pressure, which are know to impact cybersecurity behavior [51].

A second limitation concerns sampling where this research included participants studying, or working at, a university. As such, the results are representative of that population and any inference beyond that population should be avoided. On this topic, recent research argues that there are indeed demographic differences in the ability to detect phishing [52]. The number of participants is a further limitation and a higher participant number would have been preferable. In this case, data collection was stopped following a cybersecurity incident that prompted the IT department to broadcast a phishing warning. Participants performing the experiment after that event would have been exposed to information not presented to other participants and that would have introduced bias into the dataset.

A third possible discussion under the umbrella of limitations is how the different types of training were presented to the participants. The participants placed in the game group were asked to play a game before arriving for the experiment while participants in the CBMT group were subjected to training on arrival. There is, therefore, a chance that participants in the game group forgot some of the training, or forgot to play the game entirely. The design is argued to mimic the natural behavior of the two training types and both retention and failure to play are two previously discussed obstacles with game-based training delivered in a format that requires active participation [28]. Consequently, any effect of the experimental design mimics an expected effect in a natural environment.

5.2. Future Work

While training can undoubtedly support users to identify phishing emails, this study suggests that training alone is not enough and that opens up several future research directions. First, future studies could focus on combining training with modifying the way emails are presented to users. One could imagine that finding ways to make it easier for users to find and interpret phishing identifiers could improve users' ability to identify malicious emails. A possible example could be to rewrite links in the text body of emails to always show the full link address, which is unclickable, instead of allowing clickable hyperlinks with arbitrary display names. A similar possible direction is to further research predicting user susceptibility to phishing using artificial intelligence [53]. That could identify a user in need to training and then provide tailored training. A second direction for future work could be to replicate this study with a different population. That would allow for identification of differences and similarities between, for instance, technical and non-technical users, male and female users, and users of different age.

A more theoretical direction for future work could be to evaluate the strength of the relationships in the KAB model and to evaluate the relationship between behavior and actual outcomes of that behavior. In certain situations, including phishing, applying a correct behavior is not enough, since a user also has to interpret the result of that behavior. For instance, a correct behavior would make a user control the real target of a link, and to make a decision about the email the user needs to interpret the trustworthiness of the link target. Furthermore, one could assess the possible effect of usability on the relationship between the constructs in the KAB model. One can imagine that knowledge about a certain behavior is more likely to result in that behavior if the effort to comply is low.

Author Contributions: Conceptualization, J.K., M.N. and J.R.; methodology, J.K, M.N. and J.R.; software, R.R. and A.H.; validation, All.; formal analysis, J.K.; investigation, J.K.; resources, J.K.; data curation, J.K.; writing—original draft preparation, J.K.; writing—review and editing, M.N., J.R. and S.F.; supervision, S.F.; project administration, J.K.; funding acquisition, J.K, M.N., J.R., R.R. and A.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by The Swedish Post and Telecom Authority grant number 19-10617.

Institutional Review Board Statement: Ethical review and approval were waived for this study, due to fact that it does not require ethical clearance under the Swedish Ethical Review Act. Ethical Review Act dictates that research including sensitive personal data, physical interventions on living or deceased persons, methods that aim to affect persons physically och mentally, methods that can harm persons physically or mentally, or biological material from living of deceased persons [54]. Since this research does not fall under any of those criteria, ethical clearance has not been applied for. The study has been discussed with the chairperson of the council of research ethics at the University of Skövde.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Data supporting this research can be found at: https://doi.org/10.5878/g6d9-7210 (accessed on 6 March 2022).

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Abbreviation

The following abbreviations are used in this manuscript:

ISAT	Information Security Awareness Training
CBMT	Context-Based Micro-Training
SETA	Security Education, Training, and Awareness
KAB	Knowledge, Attitude, and Behaviour
SPSS	Statistical Package for the Social Sciences
ANOVA	Analysis of variance

References

1. OECD. *Hows Life in the Digital Age?* OECD Publishing: Paris, France, 2019; p. 172.
2. Owusu-Agyei, S.; Okafor, G.; Chijoke-Mgbame, A.M.; Ohalehi, P.; Hasan, F. Internet adoption and financial development in sub-Saharan Africa. *Technol. Forecast. Soc. Chang.* **2020**, *161*, 120293. [CrossRef]
3. Anderson, M.; Perrin, A. *Technology Use among Seniors*; Pew Research Center for Internet & Technology: Washington, DC, USA, 2017.
4. Bergström, A. Digital equality and the uptake of digital applications among seniors of different age. *Nord. Rev.* **2017**, *38*, 79. [CrossRef]
5. Milana, M.; Hodge, S.; Holford, J.; Waller, R.; Webb, S. A Year of COVID-19 Pandemic: Exposing the Fragility of Education and Digital in/Equalities. 2021. Available online: https://www.tandfonline.com/doi/full/10.1080/02601370.2021.1912946 (accessed on 6 March 2022)
6. Watts, G. COVID-19 and the digital divide in the UK. *Lancet Digit. Health* **2020**, *2*, e395–e396. [PubMed] [CrossRef]
7. Joseph, D.P.; Norman, J. An analysis of digital forensics in cyber security. In *First International Conference on Artificial Intelligence and Cognitive Computing*; Springer: Berlin/Heidelberg, Germany, 2019; pp. 701–708.
8. Sfakianakis, A.; Douligeris, C.; Marinos, L.; Lourenço, M.; Raghimi, O. *ENISA Threat Landscape Report 2018: 15 Top Cyberthreats and Trends*; ENISA: Athens, Greece, 2019.
9. Bhardwaj, A.; Sapra, V.; Kumar, A.; Kumar, N.; Arthi, S. Why is phishing still successful? *Comput. Fraud. Secur.* **2020**, *2020*, 15–19. [CrossRef]
10. Dark Reading. Phishing Remains the Most Common Cause of Data Breaches, Survey Says. Available online: https://www.darkreading.com/edge-threat-monitor/phishing-remains-the-most-common-cause-of-data-breaches-survey-says (accessed on 1 December 2021).
11. Butnaru, A.; Mylonas, A.; Pitropakis, N. Towards lightweight url-based phishing detection. *Future Internet* **2021**, *13*, 154. [CrossRef]
12. Gupta, B.B.; Arachchilage, N.A.; Psannis, K.E. Defending against phishing attacks: taxonomy of methods, current issues and future directions. *Telecommun. Syst.* **2018**, *67*, 247–267. [CrossRef]
13. Vishwanath, A.; Herath, T.; Chen, R.; Wang, J.; Rao, H.R. Why do people get phished? Testing individual differences in phishing vulnerability within an integrated, information processing model. *Decis. Support Syst.* **2011**, *51*, 576–586. [CrossRef]
14. Steer, J. Defending against spear-phishing. *Comput. Fraud. Secur.* **2017**, *2017*, 18–20. [CrossRef]

15. Lacey, D.; Salmon, P.; Glancy, P. Taking the bait: a systems analysis of phishing attacks. *Procedia Manuf.* **2015**, *3*, 1109–1116. [CrossRef]
16. Khan, B.; Alghathbar, K.S.; Nabi, S.I.; Khan, M.K. Effectiveness of information security awareness methods based on psychological theories. *Afr. J. Bus. Manag.* **2011**, *5*, 10862–10868.
17. Parsons, K.; McCormac, A.; Butavicius, M.; Pattinson, M.; Jerram, C. Determining employee awareness using the human aspects of information security questionnaire (HAIS-Q). *Comput. Secur.* **2014**, *42*, 165–176. [CrossRef]
18. Puhakainen, P.; Siponen, M. Improving employees' compliance through information systems security training: an action research study. *MIS Q.* **2010**, *34*, 757–778. [CrossRef]
19. Bin Othman Mustafa, M.S.; Kabir, M.N.; Ernawan, F.; Jing, W. An enhanced model for increasing awareness of vocational students against phishing attacks. In Proceedings of the 2019 IEEE International Conference on Automatic Control and Intelligent Systems (I2CACIS), Selangor, Malaysia, 29 June 2019; pp. 10–14.
20. Bada, M.; Sasse, A.M.; Nurse, J.R. Cyber security awareness campaigns: Why do they fail to change behaviour? *arXiv* **2019**, arXiv:1901.02672.
21. Reinheimer, B.; Aldag, L.; Mayer, P.; Mossano, M.; Duezguen, R.; Lofthouse, B.; von Landesberger, T.; Volkamer, M. An investigation of phishing awareness and education over time: When and how to best remind users. In Proceedings of the Sixteenth Symposium on Usable Privacy and Security (SOUPS 2020), Santa Clara, CA, USA, 7–11 August 2020; pp. 259–284.
22. Lastdrager, E.; Gallardo, I.C.; Hartel, P.; Junger, M. How Effective is Anti-Phishing Training for Children? In Proceedings of the Thirteenth Symposium on Usable Privacy and Security (SOUPS 2017), Santa Clara, CA, USA, 12–14 July 2017; pp. 229–239.
23. Junglemap. Nanolearning. Available online: https://junglemap.com/nanolearning (accessed on 7 January 2021).
24. Gokul, C.J.; Pandit, S.; Vaddepalli, S.; Tupsamudre, H.; Banahatti, V.; Lodha, S. PHISHY—A Serious Game to Train Enterprise Users on Phishing Awareness. In Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts, Melbourne, Australia, 28–31 October 2018; pp. 169–181.
25. Lim, I.K.; Park, Y.G.; Lee, J.K. Design of Security Training System for Individual Users. *Wirel. Pers. Commun.* **2016**, *90*, 1105–1120. [CrossRef]
26. Hatfield, J.M. Social engineering in cybersecurity: The evolution of a concept. *Comput. Secur.* **2018**, *73*, 102–113. [CrossRef]
27. Renaud, K.; Zimmermann, V. Ethical guidelines for nudging in information security & privacy. *Int. J. Hum.-Comput. Stud.* **2018**, *120*, 22–35.
28. Gjertsen, E.G.B.; Gjaere, E.A.; Bartnes, M.; Flores, W.R. Gamification of Information Security Awareness and Training. In Proceedings of the 3rd International Conference on Information Systems Security and Privacy, SiTePress, Setúbal, Portugal, 19–21 February 2017; pp. 59–70.
29. Abraham, S.; Chengalur-Smith, I. Evaluating the effectiveness of learner controlled information security training. *Comput. Secur.* **2019**, *87*, 101586. [CrossRef]
30. Siponen, M.; Baskerville, R.L. Intervention effect rates as a path to research relevance: information systems security example. *J. Assoc. Inf. Syst.* **2018**, *19*. [CrossRef]
31. Wen, Z.A.; Lin, Z.; Chen, R.; Andersen, E. What. hack: Engaging anti-phishing training through a role-playing phishing simulation game. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, Glasgow, UK, 4–9 May 2019; pp. 1–12.
32. Kävrestad, J.; Nohlberg, M. Assisting Users to Create Stronger Passwords Using ContextBased MicroTraining. In *IFIP International Conference on ICT Systems Security and Privacy Protection*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 95–108.
33. Siponen, M.T. A conceptual foundation for organizational information security awareness. *Inf. Manag. Comput. Secur.* **2000**, *8*, 31–41. [CrossRef]
34. Bulgurcu, B.; Cavusoglu, H.; Benbasat, I. Information security policy compliance: an empirical study of rationality-based beliefs and information security awareness. *MIS Q.* **2010**, *34*, 523–548. [CrossRef]
35. Hu, S.; Hsu, C.; Zhou, Z. Security education, training, and awareness programs: Literature review. *J. Comput. Inf. Syst.* **2021**, 1–13. [CrossRef]
36. Aldawood, H.; Skinner, G. An academic review of current industrial and commercial cyber security social engineering solutions. In Proceedings of the 3rd International Conference on Cryptography, Security and Privacy, Kuala Lumpur, Malaysia, 19–21 January 2019; pp. 110–115.
37. Al-Daeef, M.M.; Basir, N.; Saudi, M.M. Security awareness training: A review. *Proc. World Congr. Eng.* **2017**, *1*, 5–7.
38. EC-Council. The Top Types of Cybersecurity Attacks of 2019, Till Date, 2019. Available online: https://blog.eccouncil.org/the-top-types-of-cybersecurity-attacks-of-2019-till-date/ (accessed on 31 May 2021).
39. Cybint. 15 Alarming Cyber Security Facts and Stats. 2020. Available online: https://www.cybintsolutions.com/cyber-security-facts-stats/ (accessed on 20 March 2022)
40. Sharif, K.H.; Ameen, S.Y. A review of security awareness approaches with special emphasis on gamification. In Proceedings of the 2020 International Conference on Advanced Science and Engineering (ICOASE), Duhok, Iraq, 23–24 December 2020; pp. 151–156.
41. Williams, E.J.; Hinds, J.; Joinson, A.N. Exploring susceptibility to phishing in the workplace. *Int. J.-Hum.-Comput. Stud.* **2018**, *120*, 1–13. [CrossRef]

42. Chiew, K.L.; Yong, K.S.C.; Tan, C.L. A survey of phishing attacks: Their types, vectors and technical approaches. *Expert Syst. Appl.* **2018**, *106*, 1–20. [CrossRef]
43. Microsoft. Protect Yourself from Phishing. Available online: https://support.microsoft.com/en-us/windows/protect-yourself-from-phishing-0c7ea947-ba98-3bd9-7184-430e1f860a44 (accessed on 30 December 2021).
44. Imperva. Phishing Attacks. Available online: https://www.imperva.com/learn/application-security/phishing-attack-scam/ (accessed on 30 December 2021).
45. Cuve, H.C.; Stojanov, J.; Roberts-Gaal, X.; Catmur, C.; Bird, G. Validation of Gazepoint low-cost eye-tracking and psychophysiology bundle. *Behav. Res. Methods* **2021**, 1–23. [CrossRef] [CrossRef]
46. MacFarland, T.W.; Yates, J.M. Kruskal–Wallis H-test for oneway analysis of variance (ANOVA) by ranks. In *Introduction to Nonparametric Statistics for the Biological Sciences Using R*; Springer: Berlin/Heidelberg, Germany, 2016; pp. 177–211.
47. Zimmermann, V.; Renaud, K. The nudge puzzle: matching nudge interventions to cybersecurity decisions. *ACM Trans. Comput.-Hum. Interact. (TOCHI)* **2021**, *28*, 1–45. [CrossRef]
48. Van Bavel, R.; Rodríguez-Priego, N.; Vila, J.; Briggs, P. Using protection motivation theory in the design of nudges to improve online security behavior. *Int. J. Hum.-Comput. Stud.* **2019**, *123*, 29–39. [CrossRef]
49. Klimburg-Witjes, N.; Wentland, A. Hacking humans? Social Engineering and the construction of the "deficient user" in cybersecurity discourses. *Sci. Technol. Hum. Values* **2021**, *46*, 1316–1339. [CrossRef]
50. Alabdan, R. Phishing attacks survey: Types, vectors, and technical approaches. *Future Internet* **2020**, *12*, 168. [CrossRef]
51. Mashiane, T.; Kritzinger, E. Identifying behavioral constructs in relation to user cybersecurity behavior. *Eurasian J. Soc. Sci.* **2021**, *9*, 98–122. [CrossRef]
52. Das, S.; Nippert-Eng, C.; Camp, L.J. Evaluating user susceptibility to phishing attacks. *Inf. Comput. Secur.* **2022**, *309*, 1–18. [CrossRef]
53. Yang, R.; Zheng, K.; Wu, B.; Li, D.; Wang, Z.; Wang, X. Predicting User Susceptibility to Phishing Based on Multidimensional Features. *Comput. Intell. Neurosci.* **2022**, *2022*, 7058972. [CrossRef] [PubMed]
54. Swedish Research Council. Good Research Practice. Available online: https://www.vr.se/english/analysis/reports/our-reports/2017-08-31-good-research-practice.html (accessed on 30 December 2021).