Simulated Stress: A Case Study of the Effects of a Simulated Phishing Campaign on Employees' Perception, Stress and Self-Efficacy

Markus Schöps, Marco Gutfleisch, Eric Wolter, and
M. Angela Sasse, *Ruhr University Bochum*

https://www.usenix.org/conference/usenixsecurity24/presentation/schöps

This paper is included in the Proceedings of the
33rd USENIX Security Symposium.

August 14–16, 2024 • Philadelphia, PA, USA

978-1-939133-44-1

Open access to the Proceedings of the
33rd USENIX Security Symposium
is sponsored by USENIX.

Simulated Stress: A Case Study of the Effects of a Simulated Phishing Campaign on Employees' Perception, Stress and Self-Efficacy

Markus Schöps
Ruhr University Bochum, Germany

Marco Gutfleisch
Ruhr University Bochum, Germany

Eric Wolter
Ruhr University Bochum, Germany

M. Angela Sasse
Ruhr University Bochum, Germany

Abstract

Many organizations are concerned about being attacked by phishing emails and buy Simulated Phishing Campaigns (SPC) to measure and reduce their employees' susceptibility to these attacks. Whilst some prior studies reported reduced click rates after SPCs, others have raised concerns that it may have undesirable side effects: causing some employees stress, and/or reducing their self-efficacy. This would be counterproductive, since stress and self-efficacy play a key role in learning and behavior change. We report the first study in which stress and self-efficacy were measured with $n = 408$ employees immediately after they *clicked* on or *reported* a simulated phishing email they received as part of an SPC in a large organization. To obtain richer data how employees experienced the SPC, we conducted semi-structured interviews with $n = 21$ employees. We find that participants who *clicked* on and *reported* simulated phishing emails generally perceived SPCs as positive and effective, even though recent research casts doubt on this effectiveness. We further find that participants who *clicked* on simulated phishing emails had significantly higher stress levels and significantly lower phishing self-efficacy than participants who *reported* them. We further discuss the impact of our findings and conclude that the effect of SPCs on the perceived stress of employees is an important relationship that needs to be investigated in future studies.

1 Introduction

Phishing is still a prominent threat in the corporate context [76]. Companies are interested in reducing the risk of falling victim to phishing attacks, so billions of dollars are spent every year by companies on security awareness measures with the aim of countering risk and meeting security regulations [26]. One of the most propagated countermeasures of the industry are so-called simulated phishing campaigns (SPC). Phishing emails are sent to the company's own employees on behalf of the company in order to measure the employees' ability to recognize phishing emails and to sensitize and train employees. In most cases, a phishing training is embedded for those who fall for the phishing simulation to improve their phishing detection. However, according to recent findings, embedded training as part of SPCs is not effective [44]. Still, there has been no empirical evaluation of how employees who *click* on or *report* simulated phishing emails perceive those campaigns, how this affects their *self-efficacy*, and how it affects their perceived *stress*. It has been shown that stress in the workplace has a negative impact on employee performance and their well-being [24,32,52,79,85], and that, overall, "daily stressors" can have similar negative effects [29]. These stressors, even if moderate, can accumulate over time and induce many negative effects on physical and mental health [9,58,82]. Furthermore, it has been shown that employees' self-efficacy in executing a security-relevant task is essential for successful secure behavior [21,22]. Since the experience of "failing" can induce stress [39], it can be assumed that the experience of failing to successfully recognize a simulated phishing email (by clicking on it) can also induce stress. Additionally, since the experience of failure can reduce perceived self-efficacy [71], we hypothesize that failing to recognize a simulated phishing email negatively affects self-efficacy.

It is therefore crucial to investigate SPCs not only based on click rates and performance results but also to consider the individual perspective of employees who *click* on or *report* simulated phishing emails, as well as their perceived stress and self-efficacy. We focus on the following questions:

RQ1: How do employees who interact with simulated phishing emails (*click* on, or *report* them) perceive simulated phishing campaigns?

RQ2: Do employees who *click* on a simulated phishing email report different levels of stress from those who *report* it?

RQ3: Do employees who *click* on simulated phishing emails assess their self-efficacy in dealing with future phishing emails differently from those who *report* them?

Regarding **RQ2** and **RQ3**, we generated two hypotheses for the statistical analysis: **(H1)** Employees that *clicked* on the simulated phishing email rate higher *stress* levels than employees that *reported* the email; **(H2)** Employees that *clicked* on the simulated phishing email rate lower *self-efficacy* levels than employees that *reported* the email.

Methodology and Findings We cooperated with a large-scale manufacturing company employing over 16,000 employees. As part of an SPC, we embedded a questionnaire and specifically recruited those employees who either *clicked* on (4,520) or *reported* (3,034) the simulated phishing email. Our final sample consisted of 408 employees, 5,4% of the total of employees who either *clicked* or *reported*. We further recruited 21 employees for qualitative interviews. We wanted to gain a better understanding of participants' perceived stress, self-efficacy, and perception of the campaign. We found that participants who fell for the simulated phishing email reported higher levels of stress and lower levels of phishing self-efficacy. Furthermore, the campaign was perceived as effective by most participants, although various negative emotions related to the campaign were reported.

Contribution (I) We are the first to investigate the effects on stress and self-efficacy in the context of SPCs. (II) Furthermore, we contribute to the thin literature on the effects of SPCs by investigating the perception of participants who *clicked* on a phishing email and those who successfully *reported* one. (III) Finally, we discuss our results, present recommendations for industry, and highlight open challenges for researchers.

2 Related Work & Background

Here we describe previous research on SPCs (2.1), stress and emotions (2.2), and self-efficacy (2.3).

2.1 Simulated Phishing Campaigns

Simulated phishing campaigns are a common industry practice [26, 44]. During these SPCs, organizations send employees simulated, authorized phishing emails for the purpose of evaluating or improving employees' detection ability via embedded training or awareness information [62]. Even though previous studies suggest the effectiveness of SPCs in reducing *click rates* [42, 43, 80], other studies show that these campaigns have no positive effect [15, 28]. Adding to this, Lain et al. [44] conducted a large-scale and long-term study in an organization and found that embedded training was not effective in making employees more resilient to phishing. On the contrary, it made employees more susceptible. Additionally, they reported negative side effects like a false sense of security. Some researchers further hypothesize negative effects on trust relationships of employees and organizations, and question the meaningfulness of *click rates* as an IT security measurement [78]. The effort and cost of running SPCs may also be much higher than organizations think [13], contradicting the idea of these campaigns being cheap and effective solutions to improve IT security.

2.2 Stress and Emotions

Stress arises, according to Lazarus' and Folkman's transactional model [46], when a situation is evaluated as a danger to one's well-being and makes excessive demands on one's resources. This creates an imbalance between external demands and one's own coping strategies [27]. The results are cognitive, emotional, behavioral, and physiological responses [17]. When this imbalance occurs in a work-related environment, we speak of work stress [23]. The list of work stress causes is long and includes role ambiguity, lack of resources, workload, and unsound organizational policies [14]. Consequences of work stress can be the detriment of the well-being of employees, with cognitive, behavioral, emotional, and physical consequences [14], and the decline in their job performance [87]. Regarding the severity of the perceived stress, even moderate stress can have negative consequences, especially over time [9, 58, 82], and "daily stress" can lead to additive, independent effects on physical and mental health [29, 82].

When it comes to learning and memory, an integral part of training in general, stress can have different influences: For one, stress before or during learning can block the retrieval of memories [69, 84]. Although there is evidence that stress before learning improves memory formation [33], this seems to be mostly the case with emotionally arousing, specially negative, memory elements, and at the cost of neutral elements [83]. When stress is induced during learning, the formation of both neutral and emotional memory elements seems to be impaired [70]. Similarly, stress seems to impair the updating of memories with new information, leading to an inflexible, routine-like behavior [77]. Regarding IT security, the term "information security stress" (ISS) describes the stress that results from enhanced security requirements [47], which can lead to employee violation of information security policies [1, 18]. This violation behavior can also be reinforced by feelings of frustration [20]. Furthermore, findings suggest a relation between high email load, stress and susceptibility to phishing emails [65], and a relation between high workload and the likelihood of employees clicking on a phishing link [38]. A low level of job stress is also associated with higher levels of information security awareness [54]. When it comes to emotions, findings show that positive emotions positively influence protection-motivated behaviors [88], and negative emotions are frequently viewed as detrimental to motivation, performance and learning [64] In the last decades, work stress has increased rapidly [24, 61], and with this, the amount of absenteeism and sick leave caused by stress-related

illnesses [24, 79, 85]. The outcome of this are great losses of productivity and high healthcare costs [32, 52, 85].

2.3 Self-Efficacy

The concept of self-efficacy, first explored by Bandura [3, 7], defines a person's belief in their capability to successfully perform a particular task. This perceived subjective belief can have a variety of effects on cognition and affect [6] and is positively related to the implementation of behavioral change [5]. This means that, in a work context, employees are more likely to try a new behavior if they believe that they can perform it successfully [67]. The creation and development of self-efficacy beliefs can have different sources: *Mastery Experience* (experiences of success), *Vicarious Experience* (seeing others perform successfully), *Verbal Persuasion* (giving feedback and encouragement), and *Affective State* (the influence of physiological and emotional situations on the estimation of the own capacities) [4]. One of these states, *stress*, can negatively influence self-efficacy [48, 74], while the increase of self-efficacy itself can act as a buffer against the negative impact of stress [51, 68]. Additionally, the experience of failure can reduce individuals' self-efficacy [71]. Regarding IT security, an ENISA meta-review of 2019 identified self-efficacy as the only human characteristic that could be linked to "correct" cyber security behaviors [22]. Self-efficacy also shows positive effects on threat avoidance behavior [2], preventive behavior [59, 63], on the likelihood of performing phishing attack prevention behavior [21], and on securing smart homes [12]. Furthermore, the development of self-efficacy has positive effects on learning [36] and help-seeking behavior [10]. When it comes to the effects of SPCs on self-efficacy, it has been hypothesized that these campaigns reduce self-efficacy by giving employees negative, unsuccessful experiences when falling for a phishing email [78].

2.4 Research Gap

To the best of our knowledge, no previous studies have investigated the effect of SPCs on perceived *stress* or *self-efficacy* of employees who click on or report simulated phishing emails. Lastly, although one study found side effects of an SPC on the sense of security of employees [44], no other research has specifically examined the general perception that employees who interact with simulated phishing emails (by *clicking* on or *reporting*) have about SPCs.

3 Methodology

Here we describe the context of the organization and the SPC 3.1, the recruitment process and data collection procedure 3.2, and how the questionnaire and the interview guide were created 3.3. Lastly, the analysis of the data 3.4, limitations 3.5, as well as ethics and data privacy 3.6 are discussed. An overview of the study procedure is presented in Figure 1.

3.1 Organization and SPC

The data collection for the study took place in a manufacturing company headquartered in a German-speaking country but employing more than 16,000 employees globally. We specifically collected data from employees who *clicked* on (4,520 in total) or *reported* (3,034 in total) the simulated phishing email. Our final sample consisted of 408 employees, 5,4% of the 7,554 employees who either *clicked* on or *reported* the simulated phishing email. The company executives and security specialists had decided to run an SPC as part of their security awareness measures. One researcher (R3) was involved in the organization and acted as a pivot point between the company and the research team. This researcher collected the data for our study after receiving approval from the company leaders, security experts, and the work council. The platform for the SPC was provided by an external security-awareness-training company. The campaign itself consisted of three different simulated phishing emails, of which every employee received one randomly. The emails were sent out globally to over 16,000 employees in January of 2023 over the course of three days. All three emails were in English and contained spelling mistakes in the sender's address, two of which were intended to look like coming from the company and one from Microsoft Support. When an employee clicked on the link in the email, a page was shown in which the phishing simulation was revealed and a short training was displayed. The employees could *report* the phishing email by clicking on the "Report Phishing" button that was recently implemented in the company. This button is an add-on for the Outlook email client which creates a one-click solution for users to report emails. After clicking on the button the email is sent to a tool where software and humans define the emails as clean, spam, or threat. Other ways of reporting the email (e.g. informing the security team) were not counted in our study. The campaign was not announced beforehand by the company.

3.2 Recruitment and Procedure

We specifically recruited employees who *clicked* on (4,520 in total) or *reported* (3,034 in total) the simulated phishing email. For the questionnaire, participants were recruited directly after either *clicking* on the link in the email or *reporting* the email. Employees who *clicked* on the link of the simulated phishing email were shown a message on the subsequent web page that invited them to give feedback about how the campaign affected them, as well as a link to the questionnaire. Due to technical restrictions, a direct email response to the employees who *reported* the email was not possible. Instead, a pop-up window was implemented that appeared once the

phishing email was *reported* via the "Report Phishing" button. The pop-up window contained a congratulations message for successfully reporting the email as well as the same invitation to fill out the questionnaire to give feedback. A scannable QR code and a link, which possible participants could copy and paste into their browser, were implemented to access the questionnaire. Participants were not compensated and there were no selection criteria. The questionnaire was presented in German for German-speaking countries and in English for all other countries and offered a clear opt-out. Employees who did not interact with the simulated phishing email, either by ignoring it or not seeing it, were not part of our sample. The final sample consisted of 408 employees.

For the interviews, participants were recruited through a message that invited them to give more extensive feedback in a 30-minute interview. The message was displayed after the questionnaire and in a post on the internal social media. Participants were also asked to share the invitation with other colleagues. The interviews were done either remotely or in person, and, in all cases, the audio was recorded to later be transcribed.

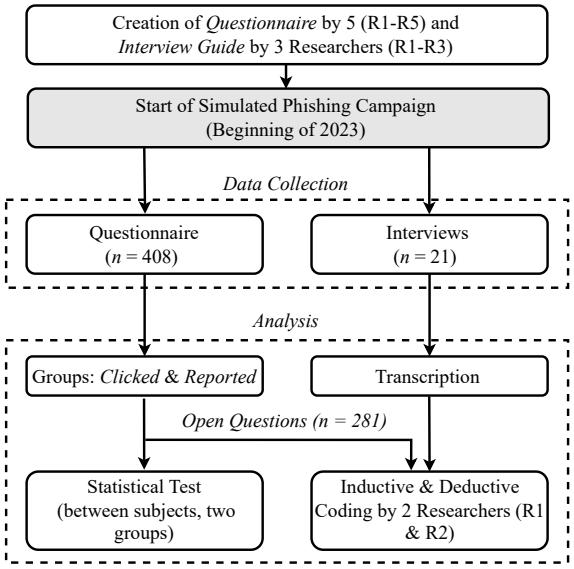

Figure 1: Overview of Study Procedure

3.3 Instrument Development

For the development of both the questionnaire and the interview guide, we were guided by our three research questions and divided both instruments accordingly. The full questionnaire, as well as the interview guide, can be found in the Appendix (see B and C).

3.3.1 Questionnaire Development

The questionnaire consisted of three parts: *Stress Measurements, Self-Efficacy Measurement, General Perception Measurements*. At the beginning of the questionnaire, participants were given the privacy statement and were informed that the questionnaire was voluntary, that no data was collected that could be traced back to them, and that the survey could be terminated at any time. Together with the company, we decided not to capture demographic data within the survey. On the one hand, the length of the questionnaire was severely limited and, on the other, we were concerned that collecting personal data would increase the drop-out rate or might have amplified a bias in our study. The length of the questionnaire was also kept short because we did not want to create any "survey fatigue".

Part I: Stress Measurements The first stress measurement that was used was the question "How stressed do you feel right now?". The question could be answered on a scale from 1 to 10 (from *Not at all* to *Fully*) and was meant to measure the direct stress after interacting with the phishing email. The second stress question was "How stressed do you feel on an average working day?", which could be answered on the same scale as the first question. This question was used to control for the possibility of one of the two groups (*Clicked* or *Reported*) being significantly more stressed on average than the other. Both of these questions were modeled after simple stress measurements like the Visual Analogue Scale (VAS) [49, 50], which shows to be as discriminating as a questionnaire when comparing the stress levels between two groups [50]. Regarding the contextualization of the stress results, Dutheil et al. [19] propose that VAS values from 5.0 to 8.2 represent an at-risk population, while a value above 8.2 represents a population in need of intervention. Other authors, like Barré et al. [8], classify values under 4.0 as low, values between 4.0 and 7.5 as moderate, and values above 7.5 as high stress. The last question was "Compared to before you clicked or reported the phishing email, how much more stressed do you feel right now?". This question could be answered on a scale from 1 to 5 (from *Significantly less stressed than before* to *Significantly more stressed than before*). Since we did not have the possibility to measure the stress before and after the interaction with the phishing email, this question was used to investigate the influence of the interaction on the perceived stress of the participants. Even though stress can be measured objectively (without self-reporting) via, for example, cortisol [57] or eye-tracking [86], these measurements present many difficulties when used in real-world contexts [31, 45, 73]. Because of this and the high number of participants, we decided to measure stress with the validated Visual Analogue Scale [49, 50].

Part II: Self-Efficacy Measurement Self-Efficacy scales should be tailored to the specific activity domain that is being investigated [6]. Because of this, we used the questionnaire developed by Ehizibue [21], which specifically measures participants' perceived self-efficacy regarding phishing, and which showed good construct reliability and validity in the past [21]. The questionnaire consisted of 4 questions (e.g. *I am confident of recognizing a phishing email*) that could be answered on a scale from 1 to 5 (from *Strongly disagree* to *Strongly agree*). This questionnaire was used to compare the self-efficacy between the two groups (*Clicked* and *Reported*). In Ehizibue's sample [21], the mean self-efficacy value was 4.51.

Part III: General Perception Measurements In this part of the questionnaire, participants were first asked to rate the *perceived importance* of the following topics: *IT security in general*, *Your role when it comes to IT security in general*, *Your role when it comes to preventing phishing attacks*, and *IT security training*. These four questions could be answered on a scale from 1 to 5 (from *Not important at all* to *Extremely important*). After this, participants were asked about their *perceived effectiveness* of the following measures to prevent clicking on phishing emails: *online phishing training*, *simulated phishing emails sent from the company*, and *technical measures (eg. email filters)*. These questions, too, could be answered on a scale from 1 to 5 (from *Not effective at all* to *Extremely effective*). Lastly, two open questions were asked: *Why do you think simulated phishing emails sent by the company are effective or not?* and *What changes would you like to see regarding the anti-phishing campaign?*. In summary, the questions in this part of the questionnaire aimed at investigating participants' perceptions of IT security measures, the phishing campaign, and their effectiveness, and how these answers differed between the two groups (*Clicked* and *Reported*).

3.3.2 Interview Guide Development

The interview guide was developed by three researchers (R1-R3) following the steps recommended by Kallio et al.'s [40] framework for developing semi-structured interview guides. The first step was to decide whether this method was appropriate to answer the research questions. Since we wanted to get a deeper understanding of the employee's diverse perceptions and possibly emotional opinions [40], we deemed this method suitable. After this, we reviewed related literature (see section 2) and developed the interview guide, which consisted of three themes related to our research questions: *General Perception*, *Stress and Emotions*, and *Self-Efficacy*. We arranged these themes and their questions in such a way, as to protect the logic and fluency of the conversation, adding follow-up questions where necessary. Next, we piloted the interview guide, which happened once with a member of our university and three times with employees of the company. The interview guide was improved after the first interview, but since no changes emerged from the three other pilotings, pilot interviews 2, 3 and 4 were included in the dataset and analysis.

The interviews themselves started with a short onboarding, in which the participants were informed about the purpose of the study, the process of data collection and storage, and the possibility of terminating the interview or skipping questions at any time. Participants then gave informed consent to record the interview. After this, a question block pertaining to the *General Perception* theme asked about how they perceived the campaign and how it was communicated. Next followed several questions asking about the *Stress and Emotions* of the participants regarding the SPC. Participants were asked to put themselves back into the situation when they first heard about the campaign and received a simulated phishing email, adding questions about what they thought, if they were stressed, and if they felt other emotions. Following this, a series of questions about the *Self-Efficacy* of the participants regarding phishing detection were asked. These questions were partly based on the sources of self-efficacy by Bandura [4] and asked, among other questions, if they had successfully *reported* a phishing email (Mastery Experience), if they had seen others successfully recognize phishing emails (Vicarious Experience), and if the company gave them feedback and encouragement (Social Persuasion). The fourth source, Affective State, was answered by the questions in the *Stress and Emotions* part of the questionnaire. Lastly, participants were once again inquired about their *General Perception* of the campaign, asking how it affected their work routine, their personal opinion about it and its effectiveness, and what other measures or changes they would like to see. Finally, we asked participants whether they would like to add anything, which has not been covered and offered a debriefing.

3.4 Analysis

The data analysis consisted of the statistical analysis of the quantitative questionnaire responses 3.4.1, and the qualitative content analysis of the interviews and the open-ended questions of the questionnaire 3.4.2.

3.4.1 Statistical Analysis

Statistical data analysis was performed using the statistical software *R* [60] and *RStudio* [66]. We used the Mann-Whitney U test to compare the means between the two groups *Clicked* and *Reported* to look for significant differences. The Mann-Whitney U test is a non-parametric test for two independent samples, that requires an ordinally or continuously scaled y-variable and doesn't require a normally distributed y-variable within the groups [55]. The Mann-Whitney U test was also

performed for the continuous data of the Self-Efficacy scale since a Shapiro-Wilk-Test showed no normal distribution of the data. We used Bonferroni-Holm corrections to correct for multiple testing and calculated effect sizes for significant comparisons, regarding a 0.1 to 0.3 as a small effect, 0.3 to 0.5 as a medium effect, and > 0.5 as a strong effect [16]. The significance level was uniformly set at < 0.05. An a priori power analysis was conducted using G*Power version 3.1.9.7 [53] to determine the minimum sample size required to test the study hypotheses. Results indicated the required sample size to achieve 95% power for detecting a medium effect, at a significance level of .05, was $n = 176$ for the Mann-Whitney U test. Thus, the obtained sample size of $n = 408$ is adequate to test the study hypotheses.

3.4.2 Qualitative Analysis

Qualitative analysis was performed using MAXQDA [75]. We used Kuckartz's qualitative content analysis [41] to evaluate the data by building categories and coding deductively and inductively. The coding of both the interviews and the open questions from the questionnaire was done by two experienced coders (R1 & R2) and was performed in multiple steps: **(I)** R1 and R2 developed preliminary codebooks deductively (one for the *Interviews* and one for the *Open Questions*) based on the research questions and the interview guide. **(II)** R1 coded all 21 interviews and 281 open questions once, in order to derive relevant themes, adding new categories inductively to the codebook. **(III)** R1 and R2 discussed the emerged codes and underlying themes, differentiating main- and sub-categories and improving the codebook with these newly created codes. **(IV)** R1 and R2 coded 11 Interviews and all Open Questions together in joint sessions, in order to further improve the codebook, refining code categories and definitions until saturation was achieved. **(V)** Using the codebook R1 and R2 coded then all previously coded and remaining *Interviews* and *Open Questions* together in joint sessions, resolving conflicts by discussion as they emerged.

Because of this method of resolving all coding issues between the coders as they appeared (which is a usual approach in our field of study [30, 35, 37, 81]) it is possible to reach a hypothetical *full* final agreement. To improve data exploration and facilitate the continuity of the process [11] memos were used and discussed during all steps of the analysis process.

3.5 Limitations

As with most studies, our study presents several limitations. First and foremost, our sample only consisted of participants who *clicked* on or *reported* the simulated phishing email via the "Report Phishing" button. Employees who did not interact with the email (either by consciously ignoring it or not seeing it) or who reported the email in other ways were not investigated in this study. Additionally, our final sample (those who answered the questionnaire, 408 in total) consisted of a small percentage (5,4%) of the number of employees we recruited from (those who either *clicked* on or *reported* the email, 7,554 in total), which limits claims about representativity. The case study was performed in one company, so the results may vary in other contexts. Additionally, we only measured the direct effects of the simulated phishing email on participants, and did not measure long-term effects of continuous exposure to these emails. The quantitative design posed challenges, since a before and after measurement was not possible. This could have led to the fact that the differences in some variables could not be attributed completely to the effect of the campaign. Due to the nature of qualitative research, the findings belonging to the interviews cannot be generalized. Also, social desirability may have influenced the answers of participants, especially when it comes to the perception of the campaign. The presentation of the questionnaire after the phishing email interaction could have led to participants thinking that this was still part of the "test", and, therefore, not responding.

3.6 Ethics & Data Privacy

Our institution did not have an institutional review board nor an ethics review board at the time of conducting the study. We adhered to the EU data privacy guidelines (GDPR) and informed all participants about their rights prior to participation. All participants gave their consent. The questionnaire collected no personal information and was, therefore, completely anonymous. The interviews were treated with confidentiality and, during transcription, all words that could identify participants or the company were removed. After transcribing the interviews, the audio files were deleted. We closely collaborated with the working council and the head of the security department to address ethical concerns. The study was then approved by the work council and the data protection teams of the countries in which the company is active. The company only received anonymized data from the study.

4 Results

Here we report the quantitative results of the questionnaire 4.1, the qualitative results of the questionnaire (open questions) 4.2, and the qualitative results of the interviews 4.3. Regarding the results of the SPC itself, 4,520 employees (26,8%) *clicked* and 3,034 (18,0%) *reported* the email. The rest did not interact with the email that they received.

4.1 Quantitative Questionnaire Results

In total, $n = 408$ participants completed the questionnaire. Of these, $n_1 = 99$ reported the phishing email (here called *reported*), and $n_2 = 309$ clicked on the link in the phishing email (here called *clicked*). On average, participants took 6.3

minutes to fill out the questionnaire. Cronbach-Alpha values showed very good reliability for the Self-Efficacy Scale (4 items; $\alpha = .86$) [72]. All reported p values are after applying Bonferroni-Holm corrections. Table 1 shows the results of the group comparisons, for which the abbreviations are written in quotes in the text below. Figure 2 in the appendix shows Boxplots for the results of the three stress measurements and the self-efficacy scale.

Table 1: Scale Response Ranges, Mean scores and p values from the group comparisons (*Clicked* and *Reported*) using Mann-Whitney U tests with Bonferroni-Holm corrections.

Variable (Scale Response Range)	Group Mean		p
	Clicked (n=309)	*Reported* (n=99)	
Stress Now (1-10)	5.40	3.39	<.001 ***
Stress Average (1-10)	5.50	4.81	.050
Stress Compared (1-5)	0.25	-0.23	<.001 ***
Self-Efficacy (Phishing) (1-5)	3.56	4.1	<.001 ***
Perceived Importance of... (all 1-5)			
ITS in General	4.56	4.7	.261
Role ITS	4.13	4.1	1.00
Role Phishing	4.3	4.39	.948
ITS Training	4.26	3.96	.015 *
Perceived Effectivity of... (all 1-5)			
Online Phishing Trainings	3.8	3.84	1.00
Phishing Campaigns	3.93	4.22	.019 *
Technical ITS Measures	4.3	4.21	1.00

Note: p = significance value, *p < .05., **p < .01., ***p < .001.

Stress-related Results The Mann-Whitney U test showed that, directly after either *clicking* on or *reporting* the simulated phishing email, the participants who *clicked* (M=5.40) had a significantly higher level of perceived stress ("Stress Now") than participants who *reported* the email (M=3.39)(U=9007.5; z=-5.824; p<.001; r=0.288). This effect was small. Regarding participants' level of stress on an average working day ("Stress Average"), there was no significant difference between the groups *clicked* (M=5.50) and *reported* (M=4.81) (U=12633; z=-1.959; p=.050). When comparing their own perceived level of stress to before clicking or reporting ("Stress Compared"), participants who *clicked* (M=0.25) indicated feeling more stressed, while participants who *reported* (M=-0.23) indicated feeling less stressed. This difference was significant (U=10771; z=-4.156; p<.001; r=0.206) with a small effect.

Self-Efficacy Results Mann-Whitney U results showed that, regarding participants' self-efficacy for detecting phishing ("Self-Efficacy (Phishing)"), participants who *clicked* (M=3.56) had a significantly lower level than participants who *reported* the email (M=4.1)(U=21810; z=-6.099; p<.001; r=0.302). This effect was medium.

General Perception Results Regarding their perceived importance of IT security in general ("ITS in General"), there were no significant differences between participants who clicked, and those who did not *click* (M=4.56) and *reported* (M=4.7) (U=16968; z=-1.124; p=.261). The same applied to participants' perceived importance of their own role regarding IT security in general ("Role ITS") (M *clicked*=4.13, M *reported*=4.1; U=15036; z=0; p=1.00) and the perceived importance of their own role regarding the prevention of phishing attacks ("Role Phishing") (M *clicked*=4.3, M *reported*=4.39; U=16391; z=-0.064; p=.948). Regarding the perceived importance of IT security training ("ITS Training"), however, participants who *clicked* (M=4.26) showed significantly higher scores than participants who *reported* (M=3.96) (U=12366; z=-2.437; p=.015; r=0.121). This effect was small. When asked about their perceived effectiveness of online phishing training, the two groups showed no significant difference (M *clicked*=3.8, M *reported*=3.84; U=15828; z=0; p=1.00). The perceived effectiveness of SPCs differed significantly between the groups, with participants who *clicked* (M=3.93) showing lower scores than participants who *reported* (M=4.22)(U=18176; z=-2.354; p=.019; r=0.117). This effect was small. The group differences regarding the perceived effectiveness of technical It measures were not significant (M *clicked*=4.3, M *reported*=4.21; U=14539; z=0; p=1.0).

Summary Participants who *clicked* reported significantly higher levels of perceived stress and lower phishing self-efficacy than participants who *reported*. Also, participants who *clicked* considered IT security training more important but SPCs less important.

4.2 Qualitative Questionnaire Results

In total, $n = 281$ participants answered at least one of the open questions from the questionnaire. Of these, $n = 13$ were excluded (if the meaning of the response could not be deciphered), resulting in a final count of $n = 268$. Of this count, $n = 196$ *clicked* and $n = 72$ *reported* the phishing email. We report counts (n_C meaning the counts of participants who *clicked*, and n_R those who *reported*), as well as what percentage of employees of the relevant group (*clicked* or *reported*) reported this. Some participants gave more than one answer, resulting in multiple codes, and some participants only answered one of the two questions. The final codebook can be found in the replication package.

4.2.1 Perception of Campaign Effectiveness

In general, $n = 224$ participants (83,6%) rated the campaign as being effective. Of the participants who *clicked* on the email, $n = 161$ (82,1%) rated the campaign as being *effective*, and of the participants who *reported* the email, $n = 63$ (87,5%). The most frequent reason given for perceiving it as effective was, in both groups, that it improved employees'

awareness (n_C = 87; *44,4%* / n_R = 38; *52,8%*). The fact that they were good *training or learning* (n_C = 25; *12,8%* / n_R = 12; *16,7%*), *real practice* (n_C = 16; *8,2%* / n_R = 15; *20,8%*), and a good *measurement* (n_C = 10; *5,1%* / n_R = 8; *11,1%*) was mentioned by participants of both groups. Next, the fact that it was a *surprise or shock* was also mentioned by participants of both groups (n_C = 9; *4,6%* / n_R = 3; *4,2%*). Additionally, some of the participants of the group that *clicked* mentioned *feedback* ($n = 4$; *2,0%*) and the *protection of the company or themselves* ($n = 10$; *5,10%*) as reasons, while some participants of the group that *reported* mentioned the feeling of *success* ($n = 2$; *2,8%*), *shame* ($n = 2$; *2,8%*), and the *communication about the topic* ($n = 2$; *2,8%*) as reasons. On the other hand, $n = 20$ participants (7,5%) rated the campaign as *not effective*. Of these, $n = 15$ (7,7%) belonged to the participants who *clicked*, and $n = 5$ (6,9%) to the participants who *reported*. Four participants mentioned the emails being *too specific* as reasons for the campaign not being effective (n_C = 3; *1,5%* / n_R = 1; *1,4%*) *"I believe phishing emails keep evolving. No way to simulate all possibilities."* – [C196], and three that they were no real *learning* (n_C = 2; *1,0%* / n_R = 1; *1,4%*). Additionally, two participants who *clicked* mentioned the creation of *mistrust* ($n = 2$; *1,0%*). Two participants who *reported* mentioned that the emails were *too easy* ($n = 2$; *2,8%*) and also two that *people talked about it* ($n = 2$; *2,8%*) *"[...] the first one informs the others, who then don't have to be "alert" anymore."* – [C196]. Three participants who *clicked*, ($n = 3$; 1,5%) mentioned reasons for and against the effectiveness of SPCs, as did three ($n = 3$; 4,2%) participants who *reported*.

4.2.2 Changes and Other Measures

Participants suggested a number of changes that they would like to see related to the SPC. The most frequently suggested change by participants who *clicked* was the wish for *more training* ($n = 24$; *12,2%*) , while only one participant who *reported* mentioned this ($n = 1$; *1,4%*). Other changes that were mentioned were that the SPCs should occur *more frequently* (n_C = 18; *9,2%* / n_R = 11; *15,3%*), that there is a need for *more information* regarding the campaign and phishing in general (n_C = 13; *6,6%* / n_R = 7; *9,7%*), the need for *technical measures* to protect against phishing (n_C = 4; *2,0%* / n_R = 7; *9,7%*) and the need for *more feedback* (n_C = 3; *1,5%* / n_R = 1; *1,4%*). Additionally, three participants who *clicked* mentioned *easier emails* ($n = 3$; *1,5%*) and two mentioned *more communication* ($n = 2$; *1,0%*) as changes, while four participants who *reported* mentioned *more difficult emails* ($n = 4$; *5,6%*) and one mentioned an *adaptation of the difficulty* to the success of the participants ($n = 1$; *1,4%*). Furthermore, one participant of each group mentioned the *termination* of the campaign as a change (n_C = 1; *0,5%* / n_R = 1; *1,4%*). Lastly, some participants mentioned that they wished for no changes (n_C = 39; *19,9%* / n_R = 11; *15,3%*).

4.2.3 Stress, Emotions and Self-Efficacy

In addition to the direct answers to the two open questions, participants also mentioned effects of the SPC on *Stress and Emotions* and *Self-Efficacy*. *Stress* was not mentioned much in either group (n_C = 3; *1,5%* / n_R = 1; *1,4%*), even though one participant who *clicked* voiced a strong negative reaction: *"[...] I now feel worthless within the company and it has now reduced me to a very negative approach to my work where i now need to switch off and calm down before i become ill."* – [C69]. *Annoyance* about the campaign was mentioned by participants in both groups (n_C = 8; *4,1%* / n_R = 2; *2,8%*): *"People who are not interested are annoyed by this and feel harassed, patronized or perhaps even deceived."* – [C191]. Two participants who *clicked* mentioned feeling *confused* ($n = 2$; *1,0%*) and also two mentioned feeling *angry* ($n = 2$; *1,0%*), while two participants who *reported* mentioned feeling *happy* ($n = 2$; *2,8%*). The negative effect on *Self-Efficacy* was only observed in statements by participants who belonged to the group that *clicked* ($n = 18$; *9,2%*): *"Before I clicked this link, I was sure I recognized phishing emails. The fact that I clicked the link showed that this is not the case."* – [C100].

Summary Most participants perceived the SPC as effective, but some wished for more training. A small number of participants mentioned negative emotions and stress, and mostly participants who *clicked* mentioned a decrease in self-efficacy because of the SPC.

4.3 Qualitative Interview Results

We first present results about the *General Perception* of the SPC 4.3.1. Next, we present results about the *Perceived Stress and other Emotions* 4.3.2, as well as the effect on participants' *Self-Efficacy* 4.3.3. Table 2 in the appendix shows the most relevant codes and their occurrences by participant. The interviews lasted between 20 and 30 minutes.

4.3.1 General Perception

Perception of Campaign All 21 participants perceived the campaign as something positive and many described that their colleagues also perceived it as something positive (P2,3,6-10,13-16,20). Regarding the reasons why they perceived it as positive, some participants stated that it was good because they had *witnessed other incidents* in the past (P1,3,8,10,14,20): *"You see in the news how many companies or authorities are shut down for weeks."* – [P14]. Others mentioned the *protection of the company* as a reason (P7,13,14,16,20,21): *"So that people deal with it, because in the end it also protects the company, if the employees pay attention to such things [...]"* – [P20]. Next, five participants mentioned that it was good because it showed them that the *organization cared about IT security* (P6,8,12,14,20). Additionally, five participants liked the *challenge* idea behind it (P3,6,7,9,20): *"Yeah, more like that, like a*

challenge, like, "I want to detect every phishing email and always report it in a timely manner."" – [P6]. Also, three participants stated that it was something that was also *useful in private* (P3,4,15), and also three that it offered a *safe space to practice* without repercussions (P1,10,12).

Perception of Campaign Effectiveness All interviewed participants perceived the campaign to be effective in some ways, mostly because they thought it increased *awareness* (P1,2,4-10,12-15,17-21). Seven participants stated that it was effective because it was *real practice* (P2,6,10,12,13,16,20): *"It trains, it's just like a little training camp [...] for reality."* – [P6]. Five participants mentioned the *shock* that it created for participants who clicked as a reason for it being effective (P2,4,7,13,21). Two participants described that the campaign was effective because it *lowered the confidence* of people who clicked (P6,11): *"[...] I personally think that you should approach such things with a not-too-strong self-confidence [...]"* – [P13]. One participant mentioned *fear* of consequences for employees who clicked as an effective characteristic of the campaign (P13): *"And maybe also, this is maybe a little bit mean, but maybe also worry that that might get to the boss, and be a little bit more careful in the future."* – [P13]. Some interviewed participants also perceived the campaign as ineffective in some ways (P7,16-19,21). Six participants (P7,8,16,18,19,21) explained that the training effect was *forgotten quickly*

Communication Most interviewed participants stated that they communicated in some way with other employees about the campaign (P1-3,6-11,14,15,17,16,21), even though half of them said that it was just briefly discussed (P2,3,6,8,11,14,15): *"I do, but not often, maybe, maybe kind of briefly mentioned in conversation or something."* – [P11]. One-third of the participants stated that they did not really communicate about the campaign (P1,4,5,12,13,16,19). Interestingly, many participants stated that there should be *no communication about the campaign* during the campaign (P3,5,8-10,13,14,18), because it could reduce the effectiveness: *"We should do it in such a way that we do not communicate it to colleagues, because then the awareness is a bit gone."* – [P3]. Only two participants said that it was positive communicating about the campaign (P2,6). Regarding the *announcement of the campaign*, four participants (P2,4,7,13) noticed that there would be a SPC, but were told about it unofficially by colleagues. The rest of the intrviewed (P3,5,6,8-21) participants stated that they noticed the campaign either after interacting with one of the emails or reading about it in the company's social media.

Changes and Other Measures More than half of the interviewed participants wished that the *frequency* of the simulated phishing emails would be increased (P1,3,6,8,10,13,14,16,18-20), even though two participants mentioned that the emails should not be sent out too often (P3,10): *"Well, I wouldn't do it too often, because then people are annoyed and then it really leads to said stress."* – [P10]. On the other hand, one participant even stated that it would be good to send out many emails in a short period of time to stress employees (P14): *"[...]that you get two or three in a week. In order to simply...then to trigger stress."* – [P14]. Many participants also mentioned that they would like more *information* about the outcome of the campaign (P1,4,7,8,13,14,16-18), which included evaluations and feedback, while others mentioned wanting more *training* (P4,6,8,11,14,18,20): *"[...] a more intensive training, maybe presence training or so."* – [P18]. Some participants described the need for more *awareness* in the company (P6,8,10,13-15,18). Regarding the *difficulty* of the emails, four participants were in favor of a higher difficulty (P5,11,17,21), while two participants wanted them to be adapted to the ability of the employees (P3,15): *"I would harass them with a different quality of phishing emails."* – [P3]. Lastly, one participant wanted the possibility to give *feedback* (P1), one more *gamification*-inspired measures (P3), and one extending the simulated-phishing to *other mediums* (P15): *"[...] like a phone social engineering campaign [...]"* – [P15].

4.3.2 Stress and Emotions

Five interviewed participants mentioned being stressed in some form or other because of the SPC (P1,2,13,15,17). Of these, one participant mentioned being stressed after receiving the email (P1): *"[...] that gut feeling: "Something's wrong, isn't it?" And that puts you in a bit of stress"* – [P1]. Three participants reported being stressed after clicking on the phishing email (P2,13,17): *"I already had a guilty conscience and the pulse was a bit higher. Because at first you don't know "Wait, was that really spam or was that a training campaign?" There was also a screen that [said that] it was a training campaign. Ehm. But that did increase my stress level significantly."* – [P17]. Two participants also mentioned being stressed because of the campaign in general (P13,15): *"Because the feeling of stress that came up for me was the fear that I work in IT and have relatively large and extensive access to some IT systems in our company. I'm always relatively quickly worried that [...] I'm the starting point of some ransomware attack or something."* – [P13]. No participant mentioned being stressed after *reporting* the email, but four participants stated seeing other colleagues being stressed after *clicking* on the email (P4,6,10,11).

When it comes to the effect of the SPC on the emotions of the participants, four interviewed participants mentioned feeling *angry* after clicking on the email (P3,12,13,20): *"The second one, I fell for it, the third one I fell for it too, and that angered me massively."* – [P3]. One participant stated feeling angry because of the timing of the campaign (P4): *"I [had a] beautiful Christmas, New Year, [...] a few days off and then you come around the corner with such a phishing campaign. On the first day, I didn't think that was quite fair."* – [P4]. One participant described feeling *shame* for *clicking* (P17), and two participants described

that other colleagues felt ashamed for *clicking* on the email (P9,10). Six participants mentioned feeling *confused* after receiving the simulated phishing email (P1,4,8,10,13,16): *"I was confused, I was very confused the moment I got that email."* – [P10]. Also, two participants experienced other colleagues being confused (P9,11). The feeling of *curiosity* was also reported by four participants (P7,8,11,18), especially when it came to how they or others performed: *"[...] curiosity. That's because I'm always interested in such topics. And I'm interested in how colleagues implement them."* – [P8]. Two participants also mentioned that they felt *relief* after *reporting* (P14,17), and two even after *clicking* (P2,13): *"I felt relief after realizing that it was just a phishing campaign and so nothing else is going on."* – [P2]. Also, three participants stated that they felt relief when they noticed that the emails, in general, were simulated and no real attack on the organization (P8,9,14): *"Being relieved, it was just a test, right? It was not a real attack on [company] now."* – [P14]. On a more positive note, eight participants described feeling happy after *reporting* the email (P1,6,7,12,14,17,19,20): *"I was rather pleased that I recognized it and also reported it, yes."* – [P6].

4.3.3 Self-Efficacy

Regarding their *Mastery Experience* (the experience of success), all but one interviewed participant stated that they successfully detected at least one simulated phishing email (P1,3-21). Of these, two participants consciously ignored the email (P10,21), while the others *reported* it (P1,3-9,11-20). Seven participants mentioned that they *clicked* on at least one simulated phishing email (P1-3,12,13,17,20). The fact that some participants mentioned both *clicking* and *reporting* is due to them recalling older SPCs. Participants also experienced the success of others (*Vicarious Experience*), with one-third of participants stating that their colleagues detected at least one of the simulated phishing emails (P2,3,5,7,8,11,13). Other participants mentioned that they experienced colleagues falling for the simulated phishing email (P6-10,14,18): *"[...] and I know that from some colleagues, that they did click twice on the links and then also felt totally insecure that moment"* – [P9]. Many participants mentioned that they did not know if other colleagues *reported* or *clicked* (P1,4,12,15-17,19-20). When it comes to the third source of self-efficacy, *Social Persuasion*, participants were asked about them getting *feedback* and *encouragement*. When asked about *feedback*, many participants mentioned the SPC itself (P1,2,4,6-8,10,12,13,16,19,21), meaning the automated email that was sent out after interacting with the simulated phishing email: *"[...] and if you do it right and press the button, then you get a pat on the back, you get a virtual one [...]"* – [P1]. The other type of feedback that was mentioned by the interviewed participants was the internal media of the company (P1,3,5,6,8,10,12,13,15,17-21), which mainly consisted of click rates reported in the internal social network or the company's intranet: *"Yes, there is always...there is always something in the intranet again afterward."* – [P3]. Three participants stated that they did not perceive receiving any kind of feedback (P9,11,14). When asked about how the company *encouraged* the participants to be successful in detecting phishing emails, many interviewed participants mentioned sources that did not come directly from the company. Many participants mentioned the motivation to *protect the company* from an incident as encouragement (P1,2,4,6,8,10,12,14,15,18): *"Because I know that if something goes wrong, the costs are immense. In the worst case, the business can't continue."* – [P15]. Others mentioned *protecting themselves* as encouragement (P1,6,8-10,14,15,20): *"[...] what motivates me is just having the fear of being the one, so being the reason [...] So to open the doors for, yes, attacks."* – [P9]. Four participants felt encouraged by the automated response after *reporting* the simulated phishing email (P6,11,18,20): *"[...] when you get this feedback: "Yes, thank you very much for reporting it, and it really was the right thing to do." I think that motivates people to remain vigilant."* – [P6]. Three participants mentioned not feeling encouraged at all (P5,17,19).

Lastly, interviewed participants were asked about their self-efficacy regarding phishing detection and the influence that the SPC had on it. Most of the participants reported that they felt somehow secure when it came to handling phishing emails (P1-7,9-16,18-21), even though some stated that being completely secure was not possible (P4,7,8,11,14,16,18,21): *"Yes, as I mentioned, so I don't think anybody is immune from that."* – [P18]. Two participants did not feel as secure (P8,17), either because of the aforementioned reason that there is no complete security, or because of failing the SPC: *"Well, I would have actually assessed myself as competent, but then I clicked on the email. That's why I wouldn't say that anymore."* – [P17]. This possible negative *influence* of the campaign on the self-efficacy regarding phishing detection was reported by four participants in total (P2,9,13,17). Eight participants stated that it *increased* their self-efficacy (P1,3,6,8,16,18,20,21): *"I would say very secure. Because I haven't misjudged any of them yet."* – [P5]. The remaining interviewed participants (P4,5,7,10-12,14-16,18-21) reported that the campaign had no real influence on them. This *increase* and *reduction* of self-efficacy as a result of the campaign is also described in Table 2 in the appendix.

Summary All interviewed participants rated the SPC as positive and effective, but some mentioned restricted communication because of it and the need for more training. Some participants mentioned feeling stressed and having their self-efficacy reduced because of the SPC. This was especially the case for participants who *clicked*.

5 Discussion

In this section, we discuss the results of the study. First, we answer our first research question (**RQ1**), by reporting that the campaign is mostly perceived as positive and effective by

participants who *clicked* on and *reported* the simulated phishing email 5.1, that participants want more training 5.2, and that the campaign, according to them, restricts communication about phishing 5.3. Next, we discuss our second research question **RQ2**, describing how the campaign increases *stress* for participants who fall for it, and confirm our first hypothesis (**H1**) that stress is higher for participants who *clicked* than those who *reported* 5.4. After that, we answer our third research question (**RQ3**), discussing the negative effect that the campaign has on the *self-efficacy* of participants who fall for it, and confirm our second hypothesis (**H2**), that participants who *clicked* on the email have lower self-efficacy than those who *reported* it 5.5. Lastly, we present *recommendations for practicioners* 5.6 and *recommendations for researchers* 5.7.

5.1 Perceived as Good and Effective

Most participants have a positive perception of the campaign (see 4.3.1 and 4.2.1). The main reasons they cited are the witnessing of other incidents and wanting to protect the company from them - which suggests that the awareness of our participants was high. Similarly, the results from the questionnaire and the interviews show that participants are motivated and care about the protection of the company, and respond positively when the company introduces ways to promote this protection 4.3.1. They perceive IT security in general, and their role regarding IT security and phishing as relatively high, with no difference between participants who *clicked* and *reported*. It seems clear that they see themselves as the main protectors of the company, as one participant stated: *"Because the human factor is one of the biggest adjusting screws that can be influenced positively or negatively. That's why it's an important topic, and it could be made even more prominent because IT security is something you can't take lightly."* – [P8].

Even though this view belongs to a sample which only represents a small percentage of the company's employees who interacted with the simulated phishing email, research shows that this view is often propagated in industry [56]: the human factor is seen as the main way to improve IT security. A possible reason for the participants' perceived effectiveness of the campaign could be that this industry view affected the company's view and this, in turn, may have "trickled down" and affected the participants' perception: that it is effective, mainly because it improves awareness. The "real practice" that participants mentioned is, in theory, a good way to establish positive IT security behavior, by promoting *mastery experience*. But this is mostly the case for participants who detect and *report* phishing emails (see section 5.5). Other reasons for the effectiveness of the campaign were that it "shocked" participants, "lowered confidence" or transmitted "fear of consequences", which can be compared to the "teachable moment" with which awareness companies think that they can transmit IT security knowledge and behavior [25], even though it is doubtful if they work [78]. This may be an explanation for why the stress that the campaign may cause is also seen as something positive by the majority of participants, even though it has mostly negative effects (see 5.4). Still, there were also doubts about the effectiveness of the campaign (see 5.2 and 5.4). Furthermore, results show that participants who *clicked* perceived the campaign as significantly less effective than participants who *reported* 4.1. This is not surprising, considering that participants who clicked "failed" the campaign. It is not clear how more frequent campaigns might influence this mostly positive perception that the participants have when the "teachable moments" also become more frequent. Besides, the fact that participants may feel the need to talk positively about their company's measures is a bias that should also be considered.

5.2 More Training and Feedback Desired

Continuing with the positive perception of the campaign, many participants wished that the emails were sent out more often, which was mostly the case for participants who *reported*. For participants who *clicked*, we can see a preference for more *training* 4.2.2 and a higher perceived effectiveness of IT security training 4.1. The perceived need for more training by the participants may be a result of the critical appraisal of their skills after failing the "test", and wanting to improve them, with the motivation to protect themselves or the company 4.3.1. More training, though, may also mean that the simulated phishing emails are not seen as enough to improve participants' IT security skills. Lastly, some participants wished for more *information* about the campaign, meaning getting feedback and seeing an evaluation of the campaign in general. Apparently, these participants wish to know the outcome of their efforts, a motivation that may be related to the "challenge" aspect of the campaign, which some participants described as positive 4.3.1. Two ways how the organization informs about the SPC and gives feedback is via click rate reporting in the internal media and short, automated feedback from the emails itself. Apparently, some participants think that this form of giving information about the SPC is too restrictive.

5.3 Restricted Communication

Restricted communication was also reported in the behavior of employees. Some participants reported that, even though communication about the campaign takes place, it is often brief. One explanation for this may be that some participants believe that there should be no communication at all 4.3.1. The "selling points" of SPCs are the "teachable moments" [25], which cannot occur when employees are forewarned. Participants who think that there should be no communication about it might have heard about these "teachable moments" and believe that talking about it is forbidden. Even though our sample is not representative of the whole organization, it

is questionable that these participants "learn" not to communicate, when communicating and warning about phishing attacks constitutes an important protection mechanism. Adding to this, since the SPC was not officially announced by the company, employees may not know that the email before them is simulated. In the worst case, employees might encounter a real phishing email but think that it's simulated, leading to them not warning their colleagues and the company.

5.4 Increased Stress when Clicked

The quantitative results showed that participants who *clicked* on the simulated phishing email reported significantly higher stress levels than those who *reported* 4.1. Therefore, we consider our first hypothesis (**H1**) confirmed. Furthermore, participants who *clicked* rated higher stress levels than before the simulated phishing email, while participants who *reported* rated lower stress levels than before. These results, which were significant, speak for the effect of the simulated phishing email on the increase/decrease of stress. The stress on an average day of both groups did not differ significantly, which supports the claim that the effect was caused by the SPC experience which immediately preceded the measurement. Stress was also reported in the qualitative results of the interviews 4.3.2. Here, especially participants who *clicked* on a simulated phishing email reported stress. Lastly, even though there were extreme reactions, stress was hardly reported in the qualitative results of the open questions 4.2.3.

When looking at the research about stress 2.2, it shows a possible explanation for why participants who *clicked* on simulated phishing reported higher stress levels: the external demands from the company to successfully detect the phishing email can be seen as too high, their own resources evaluated as too low [27, 46]. The fear of having "done something wrong" and potentially put themselves and the company in danger, may be another reason for the increased stress. Some participants mentioned feeling confused, and not only participants who *reported* the email mentioned feeling relieved, but also participants who *clicked* 4.3.2. It is possible that participants who *clicked* on simulated phishing emails thought at that moment that the attack was real, felt stressed because of it, and were relieved after knowing that they caused no harm. As mentioned in 5.1, the positive attitude regarding the phishing simulation despite the increased stress levels might be because campaigns are not done frequently, that participants think that this stress is "necessary" for the teachable moment, or that participants felt the need to talk positively about their employer.

The perceived stress level of the participants who clicked (5.40) can be categorized differently by different researchers: some categorize it as "at-risk" [19], others as "moderate" [8]. As mentioned in 2.2, stress can have negative effects on IT security behavior [1, 18, 54], and negative emotions, like the reported *anger*, 4.3.2 can reduce motivation [64, 88]. Furthermore, other findings suggest a relation between high email load, stress, and susceptibility to phishing emails [65]. This increased susceptibility is reported by Lain et al. as a side effect of SPCs [44]. It may be that stress plays a critical role in this relationship. Lastly, one important aspect of stress that needs to be mentioned is its negative effect on learning [70, 83]. Even though our results can not be generalized to the whole organization, some participants reported an increase in stress after *clicking* on the simulated phishing email. Thus, their learning of the information in the embedded training might be impaired. This effect of stress on learning could be an explanation for the low effectiveness of embedded training in SPCs that Lain et al. reported [44].

5.5 Decreased Self-Efficacy when Clicked

The results from the statistical tests showed that the participants who *clicked* on the simulated phishing email reported significantly lower phishing self-efficacy levels than participants who *reported* 4.1. Therefore, we consider our second hypothesis (**H2**) confirmed. Since we were not able to conduct a before and after test, it is not clear from the quantitative results if the lower self-efficacy regarding phishing of participants who *clicked* stems from the campaign, or if it was the reason for why these participants clicked in the first place [21]. Still, qualitative analysis shows that some participants (specially of those who *clicked*) report an effect on their phishing self-efficacy: results from the open questions show that some participants mentioned this negative effect 4.2.3, as did some participants in the interviews 4.3.3. In contrast, qualitative results from the open questions show that no participants who *reported* mentioned a decrease in phishing self-efficacy because of the campaign, while results from the interviews show that mostly participants who had never falsely clicked on a simulated phishing email mentioned an increase in phishing self-efficacy. These differences in the qualitative results are not generalizable to the whole population of employees who *clicked* or *reported*, especially because of the small sample, but they still offer a hint regarding the possible direction of the effect of simulated phishing emails on self-efficacy.

If we try to rationalize this possible decrease and increase of phishing self-efficacy, we can derive reasons from the *sources* of self-efficacy [4]: Individuals who fall for the simulated phishing email experience no success (*mastery experience*), while individuals who detect it do [71, 78]. It is not clear how often employees see the success of others (*vicarious experience*), which can foster the development of self-efficacy, in the investigated organization. Looking at the qualitative results of the interviews, two-thirds of the participants mentioned that they either saw other colleagues fail, or that they did not know how they performed 4.3.3. Furthermore, it is not clear how employees experience *social persuasion* to act successfully regarding phishing in the organization. Most participants mentioned in the interviews that *feedback* mostly

stems from the automated response of the simulated phishing email that was *clicked* on or *reported*, or by click rates reported in the internal media 4.3.3. Regarding *encouragement*, most participants mentioned other, internal sources than the company: the motivation to protect themselves or the company. Only participants who *reported* mentioned feeling encouraged by the automated response of the simulated phishing email. Lastly, participants who *clicked* might have been subjected to a negative *affective state* caused by an increase in stress [48, 74] that they reported 4.1. Concluding, there is a significant difference regarding phishing self-efficacy between participants who *clicked* on and *reported*. Even though the qualitative results from this study hint at it, it needs to be confirmed in future studies whether this difference is caused by falling for the simulated phishing email. Regardless, low self-efficacy can have negative effects not only on phishing detection [21], but also on "correct" IT security behavior [22], threat avoidance [2], and preventive behavior [59, 63].

5.6 Recommendations for Industry

Most participants in our sample believed that SPCs were effective, and other studies showed that CISOs like SPCs because they allow them to report "numbers" (*click rates*) [34]. However, this form of measurement and reporting may not be meaningful [78], and studies showed that SPCs are not an effective way of training employees [15, 28, 44]. Furthermore, the costs of implementing SPCs can be higher than anticipated [13]. We advise organizations to view these campaigns and the promises from awareness companies critically.

Our study showed that the implemented SPC significantly increased the perceived stress of participants who *clicked* on a simulated phishing email and that there is a significant difference regarding the phishing self-efficacy between participants who *clicked* on and *reported* a simulated phishing email (see Sec. 4.1, 4.3 and 4.2.3). Because of the limited representativity of our results, claims about how organizations should react to these results are limited: employees who ignored the simulated phishing email were not investigated, and their perceived stress and self-efficacy may or may not have been affected by the SPC. The perceived stress and self-efficacy of those employees who did not see the simulated phishing email were likely not affected by the SPC. Future studies are needed to further investigate these effects on stress and self-efficacy (see 5.7). Still, looking at studies which have shown negative effects of stress [70, 83] and positive effects of self-efficacy [36] on learning, we would advise organizations for any training taking place in calm situations and showing people that they can be successful.

5.7 Recommendations for Researchers

Our study showed that participants who *clicked* on a simulated phishing email felt significantly more stressed than those who *reported* it. Still, there are many opportunities for future studies to investigate stress in the context of SPCs. First, our study only gathered data from 5,4% of employees who *clicked* on or *reported* the simulated phishing email, which presents the need to investigate a larger, representative sample of participants who interact with simulated phishing emails in an organization. Second, there is a group which was not investigated in the present study: employees who did not interact with the SPC. Employees who either ignored the simulated phishing email or simply did not see it should be investigated in future studies to see how SPCs affect the perceived stress of all employees who are subjected to such campaigns. Furthermore, the perceived stress of employees should be measured long-term to see how an increased frequency of campaigns influences this stress. Moderate stress, despite its independent negative effects, can also "add up" over time, leading to negative effects on physical and mental health [9, 29, 58, 82], so it remains to be seen how repeated clicking on simulated phishing emails affects individuals.

Regarding self-efficacy, it is not clear if the significant difference between participants who *clicked* and those who *reported* was caused by the simulated phishing email itself. Future studies can confirm, via before and after measures and with a large, representative sample, if the *clicking* on a simulated phishing email negatively affects self-efficacy, as hinted at in the interviews 4.3.3. Additionally, other studies can investigate how the presence and absence of sources of self-efficacy in an organization affect the self-efficacy of employees. Measuring, with a large and representative sample, how the organization gives feedback and encouragement regarding phishing (*social persuasion*) and how employees see others being successful or not in identifying phishing emails (*vicarious experience*) can lead to valuable insights into how organizations can influence this important concept.

6 Conclusion

Our study was the first that specifically investigated stress and phishing self-efficacy as part of a real-world simulated phishing campaign of an organization. Furthermore, we investigated how participants perceived the campaign, and what emotional reactions were triggered. We specifically compared a small sample of participants who *clicked* on the campaign's phishing email and those who successfully *reported* it. We found that those who *clicked* reported a higher level of perceived stress and a lower level of phishing self-efficacy. Our findings show that SPCs can *increase stress* for individuals who click on simulated phishing emails. A possible explanation for recent research claiming that training interventions embedded in SPCs are not effective is the negative effect that stress has on learning. Future studies should investigate this effect with a representative sample and other employee groups (those who ignored or did not see the simulated phishing email).

7 Acknowledgement

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2092 CASA – 390781972.

References

[1] AGGARWAL, A., AND DHURKARI, R. K. Association between stress and information security policy non-compliance behavior: a meta-analysis. *Computers & Security 124* (2023), 102991.

[2] ARACHCHILAGE, N. A. G. User-centred security education: a game design to thwart phishing attacks. *arXiv preprint arXiv:1511.03459* (2015).

[3] BANDURA, A. Social foundations of thought and action. *Englewood Cliffs, NJ 1986*, 23-28 (1986).

[4] BANDURA, A. *Self-efficacy: The exercise of control*. W H Freeman/Times Books/ Henry Holt & Co, New York, 1997.

[5] BANDURA, A., AND ADAMS, N. E. Analysis of self-efficacy theory of behavioral change. *Cognitive therapy and research 1*, 4 (1977), 287–310.

[6] BANDURA, A., ET AL. Guide for constructing self-efficacy scales. *Self-efficacy beliefs of adolescents 5*, 1 (2006), 307–337.

[7] BANDURA, A., FREEMAN, W. H., AND LIGHTSEY, R. Self-efficacy: The exercise of control, 1999.

[8] BARRÉ, R., BRUNEL, G., BARTHET, P., AND LAURENCIN-DALICIEUX, S. The visual analogue scale: An easy and reliable way of assessing perceived stress. *Quality in Primary Health Care 1*, 1 (2017), 1–5.

[9] BEITER, R., NASH, R., MCCRADY, M., RHOADES, D., LINSCOMB, M., CLARAHAN, M., AND SAMMUT, S. The prevalence and correlates of depression, anxiety, and stress in a sample of college students. *Journal of affective disorders 173* (2015), 90–96.

[10] BERNACKI, M. L., NOKES-MALACH, T. J., AND ALEVEN, V. Examining self-efficacy during learning: Variability and relations to behavior, performance, and learning. *Metacognition and Learning 10* (2015), 99–117.

[11] BIRKS, M., CHAPMAN, Y., AND FRANCIS, K. Memoing in qualitative research: Probing data and processes. *Journal of research in nursing 13*, 1 (2008), 68–75.

[12] BORGERT, N., REITHMAIER, O. D., JANSEN, L., HILLEMANN, L., HUSSEY, I., AND ELSON, M. Home is where the smart is: Development and validation of the cybersecurity self-efficacy in smart homes (cysesh) scale. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2023), CHI '23, Association for Computing Machinery.

[13] BRUNKEN, L., BUCKMANN, A., HIELSCHER, J., AND SASSE, M. A. To do this properly, you need more resources: The hidden costs of introducing simulated phishing campaigns. In *32nd USENIX Security Symposium (USENIX Security 23)* (Anaheim, CA, Aug. 2023), USENIX Association, pp. 4105–4122.

[14] BURMAN, R., AND GOSWAMI, T. G. A systematic literature review of work stress. *International Journal of Management Studies 3*, 9 (2018), 112–132.

[15] CAPUTO, D. D., PFLEEGER, S. L., FREEMAN, J. D., AND JOHNSON, M. E. Going spear phishing: Exploring embedded training and awareness. *IEEE security & privacy 12*, 1 (2013), 28–38.

[16] COHEN, J. *Statistical power analysis for the behavioral sciences*. Academic press, 2013.

[17] COLLIGAN, T. W., AND HIGGINS, E. M. Workplace stress: Etiology and consequences. *Journal of workplace behavioral health 21*, 2 (2006), 89–97.

[18] D'ARCY, J., HERATH, T., AND SHOSS, M. K. Understanding employee responses to stressful information security requirements: A coping perspective. *Journal of management information systems 31*, 2 (2014), 285–318.

[19] DUTHEIL, F., PEREIRA, B., MOUSTAFA, F., NAUGHTON, G., LESAGE, F.-X., AND LAMBERT, C. At-risk and intervention thresholds of occupational stress using a visual analogue scale. *PLoS One 12*, 6 (2017), e0178948.

[20] D'ARCY, J., AND TEH, P.-L. Predicting employee information security policy compliance on a daily basis: The interplay of security-related stress, emotions, and neutralization. *Information & Management 56*, 7 (2019), 103151.

[21] EHIZIBUE, D. Investigation of individuals' behavior towards phishing attacks using the health belief model. B.S. thesis, University of Twente, 2022.

[22] ENISA. Cybersecurity culture guidelines: Behavioural aspects of cybersecurity. *European Union Agency for Network and Information Security (ENISA)* (2019).

[23] FLOREA, R., AND FLOREA, R. Individual and organizational implications of work-related stress. *Economy Transdisciplinarity Cognition 19*, 1 (2016), 28.

[24] FOY, T., DWYER, R. J., NAFARRETE, R., HAMMOUD, M. S. S., AND ROCKETT, P. Managing job performance, social support and work-life conflict to reduce workplace stress. *International Journal of Productivity and Performance Management 68*, 6 (2019), 1018–1041.

[25] FRANZ, A., ZIMMERMANN, V., ALBRECHT, G., HARTWIG, K., REUTER, C., BENLIAN, A., AND VOGT, J. {SoK}: Still plenty of phish in the sea—a taxonomy of {User-Oriented} phishing interventions and avenues for future research. In *Seventeenth Symposium on Usable Privacy and Security (SOUPS 2021)* (2021), pp. 339–358.

[26] GLOBENEWSWIRE. Global cybersecurity awareness training market size & trends, 2022.

[27] GOLDSTEIN, D. S., AND MCEWEN, B. Allostasis, homeostats, and the nature of stress. *Stress 5*, 1 (2002), 55–58.

[28] GORDON, W. J., WRIGHT, A., GLYNN, R. J., KADAKIA, J., MAZZONE, C., LEINBACH, E., AND LANDMAN, A. Evaluation of a mandatory phishing training program for high-risk employees at a us healthcare system. *Journal of the American Medical Informatics Association 26*, 6 (2019), 547–552.

[29] GRZYWACZ, J. G., ALMEIDA, D. M., NEUPERT, S. D., AND ETTNER, S. L. Socioeconomic status and health: A micro-level analysis o exposure and vulnerability to daily stressors. *Journal of health and social behavior 45*, 1 (2004), 1–16.

[30] GUTFLEISCH, M., KLEMMER, J. H., BUSCH, N., ACAR, Y., SASSE, M. A., AND FAHL, S. How does usable security (not) end up in software products? results from a qualitative interview study. In *2022 IEEE Symposium on Security and Privacy (SP)* (2022), IEEE, pp. 893–910.

[31] HALPERN, C. T., WHITSEL, E. A., WAGNER, B., AND HARRIS, K. M. Challenges of measuring diurnal cortisol concentrations in a large population-based field study. *Psychoneuroendocrinology 37*, 4 (2012), 499–508.

[32] HASSARD, J., TEOH, K. R., VISOCKAITE, G., DEWE, P., AND COX, T. The cost of work-related stress to society: A systematic review. *Journal of occupational health psychology 23*, 1 (2018), 1.

[33] HENCKENS, M. J., HERMANS, E. J., PU, Z., JOËLS, M., AND FERNÁNDEZ, G. Stressed memories: how acute stress affects memory formation in humans. *Journal of Neuroscience 29*, 32 (2009), 10111–10119.

[34] HIELSCHER, J., MENGES, U., PARKIN, S., KLUGE, A., AND SASSE, M. A. "Employees Who Don't Accept the Time Security Takes Are Not Aware Enough": The CISO View of Human-Centred Security. In *32st USENIX Security Symposium (USENIX Security 23)* (Boston, MA, Aug. 2023), USENIX Association.

[35] HIELSCHER, J., SCHÖPS, M., MENGES, U., GUTFLEISCH, M., HELBLING, M., AND SASSE, M. A. Lacking the tools and support to fix friction: Results from an interview study with security managers. In *Nineteenth Symposium on Usable Privacy and Security (SOUPS 2023)* (2023), pp. 131–150.

[36] HODGES, C. B. Self-efficacy in the context of online learning environments: A review of the literature and directions for research. *Performance Improvement Quarterly 20*, 3-4 (2008), 7–25.

[37] HÖLTERVENNHOFF, S., KLOSTERMEYER, P., WÖHLER, N., ACAR, Y., AND FAHL, S. {"I} wouldn't want my unsafe code to run my {pacemaker"}: An interview study on the use, comprehension, and perceived risks of unsafe rust. In *32nd USENIX Security Symposium (USENIX Security 23)* (2023), pp. 2509–2525.

[38] JALALI, M. S., BRUCKES, M., WESTMATTELMANN, D., AND SCHEWE, G. Why employees (still) click on phishing links: investigation in hospitals. *Journal of medical Internet research 22*, 1 (2020), e16775.

[39] JOHNSON, J., PANAGIOTI, M., BASS, J., RAMSEY, L., AND HARRISON, R. Resilience to emotional distress in response to failure, error or mistakes: A systematic review. *Clinical psychology review 52* (2017), 19–42.

[40] KALLIO, H., PIETILÄ, A.-M., JOHNSON, M., AND KANGASNIEMI, M. Systematic methodological review: developing a framework for a qualitative semi-structured interview guide. *Journal of advanced nursing 72*, 12 (2016), 2954–2965.

[41] KUCKARTZ, U. *Qualitative text analysis: A guide to methods, practice & using software*. SAGE, Los Angeles and London and New Delhi and Singapore and Washington, DC, 2014.

[42] KUMARAGURU, P., CRANSHAW, J., ACQUISTI, A., CRANOR, L., HONG, J., BLAIR, M. A., AND PHAM, T. School of phish: a real-world evaluation of anti-phishing training. In *Proceedings of the 5th Symposium on Usable Privacy and Security* (2009), pp. 1–12.

[43] KUMARAGURU, P., SHENG, S., ACQUISTI, A., CRANOR, L. F., AND HONG, J. Lessons from a real world evaluation of anti-phishing training. In *2008 eCrime Researchers Summit* (2008), IEEE, pp. 1–12.

[44] LAIN, D., KOSTIAINEN, K., AND ČAPKUN, S. Phishing in organizations: Findings from a large-scale and long-term study. In *2022 IEEE Symposium on Security and Privacy (SP)* (2022), pp. 842–859.

[45] LAPPI, O. Eye tracking in the wild: the good, the bad and the ugly. *Journal of Eye Movement Research 8*, 5 (2015).

[46] LAZARUS, R. S., AND FOLKMAN, S. *Stress, appraisal, and coping*. Springer publishing company, 1984.

[47] LEE, C., LEE, C. C., AND KIM, S. Understanding information security stress: Focusing on the type of information security compliance activity. *Computers & Security 59* (2016), 60–70.

[48] LEE, J., KIM, E., AND WACHHOLTZ, A. The effect of perceived stress on life satisfaction: The mediating effect of self-efficacy. *Ch'ongsonyonhak yongu 23*, 10 (2016), 29.

[49] LESAGE, F., AND BERJOT, S. Validity of occupational stress assessment using a visual analogue scale. *Occupational medicine 61*, 6 (2011), 434–436.

[50] LESAGE, F.-X., BERJOT, S., AND DESCHAMPS, F. Clinical stress assessment using a visual analogue scale. *Occupational medicine 62*, 8 (2012), 600–605.

[51] MAKARA-STUDZIŃSKA, M., GOLONKA, K., AND IZYDORCZYK, B. Self-efficacy as a moderator between stress and professional burnout in firefighters. *International journal of environmental research and public health 16*, 2 (2019), 183.

[52] MATILLA-SANTANDER, N., GONZÁLEZ-MARRÓN, A., MARTÍN-SÁNCHEZ, J. C., LIDÓN-MOYANO, C., CARTANYÀ-HUESO, À., AND MARTÍNEZ-SÁNCHEZ, J. M. Precarious employment and health-related outcomes in the european union: a cross-sectional study. *Critical Public Health 30*, 4 (2020), 429–440.

[53] MAYR, S., ERDFELDER, E., BUCHNER, A., AND FAUL, F. A short tutorial of gpower. *Tutorials in quantitative methods for psychology 3*, 2 (2007), 51–59.

[54] MCCORMAC, A., CALIC, D., PARSONS, K., BUTAVICIUS, M., PATTINSON, M., AND LILLIE, M. The effect of resilience and job stress on information security awareness. *Information & Computer Security 26*, 3 (2018), 277–289.

[55] MCKNIGHT, P. E., AND NAJAB, J. Mann-whitney u test. *The Corsini encyclopedia of psychology* (2010), 1–1.

[56] MENGES, U., HIELSCHER, J., BUCKMANN, A., KLUGE, A., SASSE, M. A., AND VERRET, I. Why it security needs therapy. In *European Symposium on Research in Computer Security* (2021), Springer, pp. 335–356.

[57] NICOLSON, N. A. Measurement of cortisol. *Handbook of physiological research methods in health psychology 1* (2008), 37–74.

[58] ÖHMAN, L., BERGDAHL, J., NYBERG, L., AND NILSSON, L.-G. Longitudinal analysis of the relation between moderate long-term stress and health. *Stress and health: Journal of the International Society for the Investigation of Stress 23*, 2 (2007), 131–138.

[59] OZDEMIR, S., NG, S., CHAUDHRY, I., AND FINKELSTEIN, E. A. Adoption of preventive behaviour strategies and public perceptions about covid-19 in singapore. *International Journal of Health Policy and Management 11*, 5 (2022), 579–591.

[60] R CORE TEAM. R: A language and environment for statistical computing, 2023.

[61] RIGÓ, M., DRAGANO, N., WAHRENDORF, M., SIEGRIST, J., AND LUNAU, T. Work stress on rise? comparative analysis of trends in work stressors using the european working conditions survey. *International Archives of Occupational and Environmental Health 94* (2021), 459–474.

[62] RIZZONI, F., MAGALINI, S., CASAROLI, A., MARI, P., DIXON, M., AND COVENTRY, L. Phishing simulation exercise in a large hospital: A case study. *Digital Health 8* (2022), 20552076221081716.

[63] ROSENSTOCK, I. M., STRECHER, V. J., AND BECKER, M. H. Social learning theory and the health belief model. *Health education quarterly 15*, 2 (1988), 175–183.

[64] ROWE, A. D., AND FITNESS, J. Understanding the role of negative emotions in adult learning and achievement: A social functional perspective. *Behavioral sciences 8*, 2 (2018), 27.

[65] ROZENTALS, E. Email load and stress impact on susceptibility to phishing and scam emails, 2021.

[66] RSTUDIO TEAM. *RStudio: Integrated Development Environment for R*. RStudio, PBC., Boston, MA, 2020.

[67] SASSE, M. A., HIELSCHER, J., FRIEDAUER, J., AND BUCKMANN, A. Rebooting it security awareness–how organisations can encourage and sustain secure behaviours. In *European Symposium on Research in Computer Security* (2022), Springer, pp. 248–265.

[68] SCHÖNFELD, P., BRAILOVSKAIA, J., BIEDA, A., ZHANG, X. C., AND MARGRAF, J. The effects of daily stress on positive and negative mental health: Mediation through self-efficacy. *International Journal of Clinical and Health Psychology 16*, 1 (2016), 1–10.

[69] SCHWABE, L., HERMANS, E. J., JOËLS, M., AND ROOZENDAAL, B. Mechanisms of memory under stress. *Neuron* (2022).

[70] SCHWABE, L., AND WOLF, O. T. Learning under stress impairs memory formation. *Neurobiology of learning and memory 93*, 2 (2010), 183–188.

[71] SMITH, S. A., KASS, S. J., ROTUNDA, R. J., AND SCHNEIDER, S. K. If at first you don't succeed: Effects of failure on general and task-specific self-efficacy and performance. *North American Journal of Psychology* (2006).

[72] STREINER, D. L. Starting at the beginning: an introduction to coefficient alpha and internal consistency. *Journal of personality assessment 80*, 1 (2003), 99–103.

[73] UTTLEY, J., SIMPSON, J., AND QASEM, H. Eye-tracking in the real world: Insights about the urban environment. In *Handbook of Research on Perception-Driven Approaches to Urban Assessment and Design*. IGI Global, 2018, pp. 368–396.

[74] VAN RAALTE, L. J., AND POSTEHER, K. A. Examining social support, self-efficacy, stress, and performance, in us division i collegiate student-athletes' academic and athletic lives. *Journal for the Study of Sports and Athletes in Education 13*, 2 (2019), 75–96.

[75] VERBI SOFTWARE. Maxqda 2020, 2023.

[76] VERIZON. 2023 data breach investigations report, 2023.

[77] VOGEL, S., AND SCHWABE, L. Learning and memory under stress: implications for the classroom. *npj Science of Learning 1*, 1 (2016), 1–10.

[78] VOLKAMER, M., SASSE, M. A., AND BOEHM, F. Analysing simulated phishing campaigns for staff. In *Computer Security: ESORICS 2020 International Workshops, DETIPS, DeSECSys, MPS, and SPOSE, Guildford, UK, September 17–18, 2020, Revised Selected Papers 25* (2020), Springer, pp. 312–328.

[79] WALLENSTEN, J., ÅSBERG, M., WIKLANDER, M., AND NAGER, A. Role of rehabilitation in chronic stress-induced exhaustion disorder: A narrative review. *Journal of rehabilitation medicine 51*, 5 (2019), 331–342.

[80] WASH, R., AND COOPER, M. M. Who provides phishing training? facts, stories, and people like me. In *Proceedings of the 2018 chi conference on human factors in computing systems* (2018), pp. 1–12.

[81] WERMKE, D., WÖHLER, N., KLEMMER, J. H., FOURNÉ, M., ACAR, Y., AND FAHL, S. Committed to trust: A qualitative study on security & trust in open source software projects. In *2022 IEEE Symposium on Security and Privacy (SP)* (2022), IEEE, pp. 1880–1896.

[82] WHEATON, B. Sampling the stress universe. In *Stress and mental health: Contemporary issues and prospects for the future*. Springer, 1994, pp. 77–114.

[83] WOLF, O. T. Stress and memory in humans: twelve years of progress? *Brain research 1293* (2009), 142–154.

[84] WOLF, O. T. Stress and memory retrieval: mechanisms and consequences. *Current Opinion in Behavioral Sciences 14* (2017), 40–46.

[85] WOLVETANG, S., VAN DONGEN, J. M., SPEKLÉ, E., COENEN, P., AND SCHAAFSMA, F. Sick leave due to stress, what are the costs for dutch employers? *Journal of Occupational Rehabilitation 32*, 4 (2022), 764–772.

[86] YOUSEFI, M. S., REISI, F., DALIRI, M. R., AND SHALCHYAN, V. Stress detection using eye tracking data: An evaluation of full parameters. *IEEE Access 10* (2022), 118941–118952.

[87] YU, J., PARK, J., AND HYUN, S. S. Impacts of the covid-19 pandemic on employees' work stress, well-being, mental health, organizational citizenship behavior, and employee-customer identification. *Journal of Hospitality Marketing & Management 30*, 5 (2021), 529–548.

[88] ZHEN, J., XIE, Z., AND DONG, K. Positive emotions and employees' protection-motivated behaviours: a moderated mediation model. *Journal of Business Economics and Management 21*, 5 (2020), 1466–1485.

A Replication Package

To make our study reproducible and allow for easy access for meta-research, we publish a replication package[1] containing the following documents:

1. **The full code books**, created as part of the analysis described in Section 3.4.2.
2. **The Questionnaire and Interview Guide**, created as part of the instrument development described in Sections 3.3.1 and 3.3.2 .
3. **Figures and Tables**, created as part of the results described in Sections 4.1 and 4.3

B Questionnaire

A.1 How stressed do you feel right now?

1 2 3 4 5 6 7 8 9 10
□ □ □ □ □ □ □ □ □ □
Not at all *Fully*

A.2 How stressed do you feel on an average working day?

1 2 3 4 5 6 7 8 9 10
□ □ □ □ □ □ □ □ □ □
Not at all *Fully*

A.3 Compared to before you clicked or reported the phishing mail, how much more stressed do you feel right now?

○ Significantly more stressed than before
○ Slightly more stressed than before
○ Just as stressed as before
○ Slightly less stressed than before
○ Significantly less stressed than before

A.4 Here you will find a list of statements on possible attitudes of employees towards phishing emails. Please indicate to what extent you agree with the following statements

- I am confident of recognizing a phishing email.
- I can recognize a phishing email even if there was no one around to help me.
- I can recognize a malicious URL from a legitimate URL.
- I am sure of the steps to follow to recognize a phishing email.

○ Strongly agree
○ Agree

[1] https://doi.org/10.6084/m9.figshare.25990963

○ Neither agree nor disagree
○ Disagree
○ Strongly disagree

A.5 How important do you consider...

- IT security in general.
- Your role when it comes to IT security in general.
- Your role when it comes to preventing phishing attacks.
- IT security training.

○ Extremely important
○ Very important
○ Moderately important
○ Slightly important
○ Not important at all

A.6 How effective do you consider...

- Online phishing trainings to prevent clicking on real phishing mails.
- Simulated phishing mails sent by the company to prevent clicking on real phishing mails.
- Technical measures (email filter, etc.) to prevent clicking on real phishing mails.

○ Extremely effective
○ Very effective
○ Moderately effective
○ Slightly effective
○ Not effective at all

A.7 Why do you think simulated phishing emails sent by the company are effective or not?

A.8 What changes would you like to see regarding the anti-phishing campaign?

C Interview Guide

B.1 General Perception (1/2)

- Can you tell me what you heard about the campaign and how it went?
  - How did you first hear about the campaign?
  - Was the campaign communicated to you in any other way?
  - Did you have any questions or uncertainties?

B.2 Stress and Emotions

- Please put yourself back into the situation in which you first learned about the campaign.
  - What were you thinking at that moment?
  - Did this situation stress you? Why?
  - Did this situation trigger any other emotions in you?
  - Did this situation bother you even longer afterwards?

B.3 Self-Efficacy

- How confident do you feel about dealing with phishing emails?
  - What influence did the campaign have on this?
  - *If not mentioned:* Have you ever successfully reported a simulated phishing email?

- Do you get to see if colleagues recognize phishing emails?
  - How do they view the campaign? Why?

- In what form does the company give you feedback on the results of the campaign?

- In what way does the company encourage you to be successful in the campaign?

B.4 General Perception (2/2)

- Do you exchange information about the campaign with anyone?
  - With whom? About what?
  - *Additional stress question:* Were the others stressed?

- At this point, what would you do if you were confronted with a potential phishing email?

- More generally, what is your personal opinion of the phishing campaign?

- Do you think that the campaign will help employees click on phishing emails less often in the future? What about you?

- Would you like to see other measures to help you click on phishing emails less often?
  - What changes would you like to see regarding the anti-phishing campaign?

D Figures & Tables

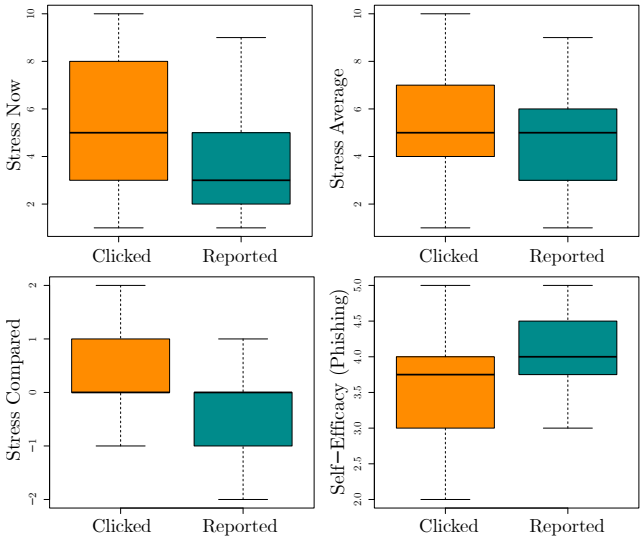

Figure 2: Boxplots of perceived stress directly after clicking/reporting ("Stress Now"), average stress, stress compared to before clicking/reporting ("Stress Compared") and phishing self-efficacy for the two Groups *Clicked* and *Reported*.

Table 2: Occurences of the most relevant codes by participants.

Code	**Participants**																					
	P1	P2	*P3*	*P4*	*P5*	*P6*	*P7*	*P8*	*P9*	*P10*	*P11*	*P12*	*P13*	*P14*	*P15*	*P16*	*P17*	*P18*	*P19*	*P20*	*P21*	*Total*
Clicked	●	●	●	○	○	○	○	○	○	○	○	●	●	○	○	○	●	○	○	●	○	7
*Reported (*Ignored)*	●	○	●	●	●	●	●	●	●	●*	●	●	●	●	●	●	●	●	●	●	●*	20
Stress	●	●	○	○	○	○	○	○	○	○	○	○	●	○	●	○	●	○	○	○	○	5
Stress Colleagues	○	○	○	●	○	●	○	○	○	●	●	○	○	○	○	○	○	○	○	○	○	4
Anger	○	○	●	●	○	○	○	○	○	○	○	●	●	○	○	○	○	○	○	●	○	5
Shame	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	●	○	○	○	○	1
Confusion	●	○	○	●	○	○	○	●	○	●	○	○	○	○	○	●	○	○	○	○	○	5
Curiosity	○	○	○	○	○	○	●	●	○	○	●	○	○	○	○	○	○	●	○	○	○	4
Relief	○	●	○	○	○	○	○	●	●	○	○	○	●	●	○	○	●	●	○	○	○	7
Happiness	●	○	○	○	○	●	●	○	○	○	○	●	○	●	○	○	●	○	●	●	○	8
Self-Efficacy Increase	●	○	●	○	○	●	○	●	○	○	○	○	○	○	○	●	○	●	○	●	●	8
Self-Efficacy Reduction	○	●	○	○	○	○	○	○	●	○	○	○	●	○	○	○	●	○	○	○	○	4