It's a Match - Enhancing the Fit between Users and Phishing Training through Personalisation

Lorin Schöni
Security, Privacy & Society
ETH Zurich
Zurich, Switzerland
lorin.schoeni@gess.ethz.ch

Neele Roch
Security, Privacy & Society
ETH Zurich
Zurich, Switzerland
neele.roch@gess.ethz.ch

Hannah Sievers
Security, Privacy & Society
ETH Zurich
Zurich, Switzerland
hannah.sievers@gess.ethz.ch

Martin Strohmeier
armasuisse
Thun, Switzerland
martin.strohmeier@armasuisse.ch

Peter Mayer
Department of Mathematics and Computer Science
University of Southern Denmark
Odense, Denmark
Institute of Applied Informatics and Formal Description Methods
Karlsruhe Institute of Technology
Karlsruhe, Germany
mayer@imada.sdu.dk

Verena Zimmermann
Department of Humanities, Social and Political Sciences
ETH Zürich
Zürich, Switzerland
verena.zimmermann@gess.ethz.ch

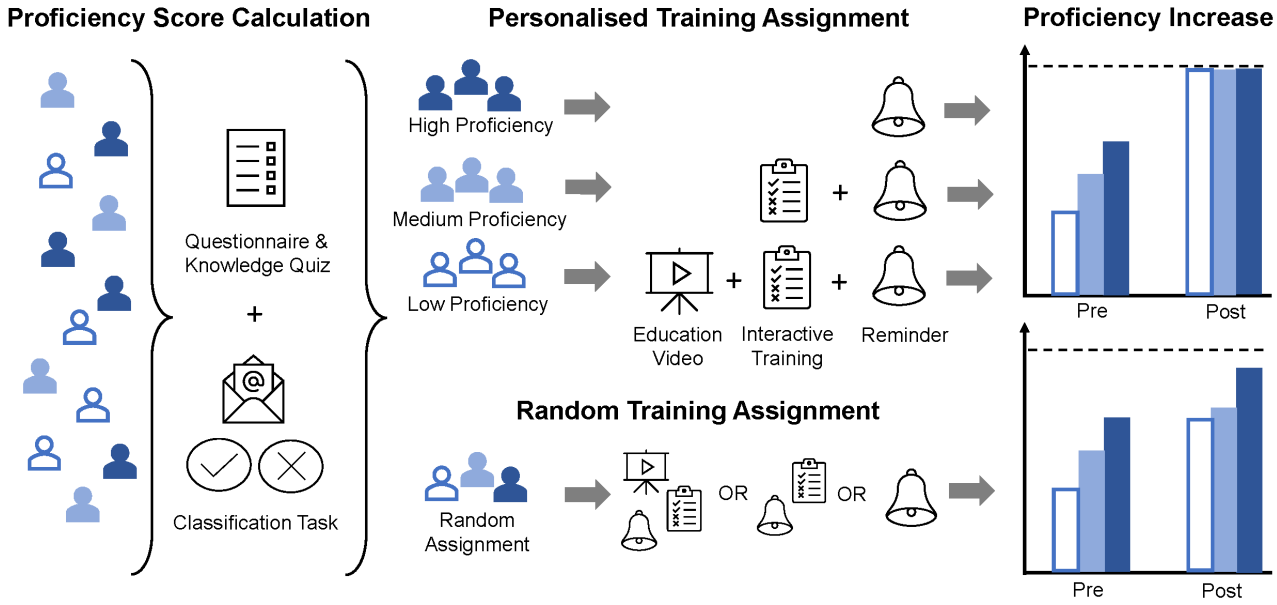


Figure 1: We categorised users in terms of their phishing proficiency based on variables like phishing knowledge and detection ability in a classification task. Then, the users were either assigned a personalised training based on their proficiency level, or randomly assigned to one of the training variants regardless of their proficiency level. Personalised training raised proficiency to a desired level, while random assignment showed less clear effects.

Abstract

Effective training is essential for enhancing users' ability to detect phishing attempts. Personalised training offers huge potential to more closely align training content with individuals' needs and skill levels. In an online study, we assigned N=342 participants to personalised training or a random training variant to compare their effectiveness. The personalisation was based on a phishing

CHI '25, Yokohama, Japan

ACM ISBN 979-8-4007-1394-1/25/04
https://doi.org/10.1145/3706598.3713845

proficiency score calculated from factors such as detection ability, knowledge, and security attitude. After training, the participants demonstrated greater proficiency, with an increased ability to detect phishing emails and higher security attitudes. These effects were most pronounced in the personalised condition, demonstrating the potential of personalisation to improve training outcomes. Overall, personalised training levelled the playing field, efficiently bringing all groups, regardless of their initial proficiency, to a comparable and desired post-training phishing proficiency level. Finally, we derived recommendations for designing personalised phishing training content and assigning users to suitable training programmes.

CCS Concepts

• **Human-centered computing** → **Empirical studies in HCI**; • **Security and privacy** → *Usability in security and privacy*; **Phishing**.

Keywords

Phishing, Personalisation, Training, Human-Centred Security

ACM Reference Format:

Lorin Schöni, Neele Roch, Hannah Sievers, Martin Strohmeier, Peter Mayer, and Verena Zimmermann. 2025. It's a Match - Enhancing the Fit between Users and Phishing Training through Personalisation. In *CHI Conference on Human Factors in Computing Systems (CHI '25), April 26–May 01, 2025, Yokohama, Japan.* ACM, New York, NY, USA, 25 pages. https://doi.org/10.1145/3706598.3713845

1 Introduction

Phishing poses a critical cybersecurity challenge, with the volume of phishing emails reaching unprecedented levels and projected to increase further [4, 75]. While technical barriers are useful at countering this threat [76], human-centred approaches like training that enables humans to detect phishing are equally important [75, 80]. Recent HCI research stressed the importance of the human factor in cybersecurity [80], reiterated once more in the CHI 2023 opening keynote [75]. Yet, phishing training generally follows a 'one size fits all' approach (e.g., [13, 18, 32, 74]). Accordingly, it neglects individual differences [1, 14, 25, 67, 71] and context [5, 15]. Related research has, however, repeatedly shown that adjusting phishing training to the target users' knowledge or experience can be highly beneficial and increase its effectiveness [1, 29]. Personalised approaches have been proposed as a key factor for training success [5], are shown to enhance the effectiveness of anti-phishing training [29], and are recommended as central success factors by cybersecurity professionals [25]. Already in other domains—most prominently in primary and secondary education [9]—personalised learning has been implemented with promising results [66]. For example, Siddique et al [57] compared a personalised approach based on learning styles, working memory capacity, and prior knowledge to traditional classroom teaching, finding that the personalised group showed a better learning performance. While the idea of personalised learning has been well-established since 2010 [9], the research field is experiencing increased attention due to emerging AI tools that can facilitate personalisation of learning content for individual learners [45, 64]. For instance, Perez-Ortiz et al. [44] proposed X5Learn, an AI-based personalised learning companion with a recommender system that adapts to users' learning preferences.

However, there are several barriers hindering the implementation of personalisation in cybersecurity training. Fundamentally, there is a lack of empirical research on *how* this personalisation should happen, i.e., *what data* is used to personalise and *to what degree* the content is modified. There are also practical limitations that can hinder the deployment of personalisation. Institutions often prioritise simple and straightforward training due to limited cybersecurity budgets, and at the same time need to comply with data privacy standards [58]. Therefore, we evaluate a phishing training that can be personalised in a straightforward way and does not rely on sensitive information (e.g., potentially discriminating demographics like age). To that end, we developed and evaluated a personalised training that consists of modular components, each selected to best fit users' current phishing proficiency and needs. Building on prior work [54], we assigned a score that weighs users' phishing-related background and proficiency to personalise training by selecting the most suitable training variant. The design of the training components and the subsequent matching process is based on the security learning curve [53], a stair-like model of successive steps. It assumes that certain steps like general awareness and understanding, or acquiring the underlying ability, are prerequisites for developing and embedding secure behaviour. Therefore, the process can account for pre-training differences and bring everyone to the same desired proficiency level with enhanced efficiency, i.e., users with an already high proficiency level might require less training to reach the desired proficiency level.

Research Aim. To address the shortage of empirical evaluations of personalised training in the phishing or cybersecurity domain, we aim to explore the potential of a personalised phishing training compared to a randomly assigned control and an education-only baseline variant. We therefore investigate five hypotheses:

- **H1:** All personalised groups show an increase in a) phishing proficiency, b) cybersecurity awareness, and c) self-estimated proficiency.
- **H2:** Participants only experiencing the education training element show lower increases in a) phishing proficiency, b) cybersecurity awareness, and c) self-estimated proficiency than participants experiencing all training elements.
- **H3:** Participants categorised into lower proficiency groups show larger increases in a) phishing proficiency, b) cybersecurity awareness, and c) self-estimated proficiency as compared to higher proficiency groups.
- **H4:** Compared to random assignment, participants assigned through personalisation show larger increases in a) phishing proficiency, b) cybersecurity awareness, and c) self-estimated proficiency.
- **H5:** Participants prefer personalisation compared to random assignment.
  - **H5a:** Participants prefer the training personalisation that corresponds to their proficiency level.

The contributions of this research are fourfold:

- First, we provide a systematic empirical evaluation of a personalised phishing training in comparison to random assignment and a 'one-size-fits-all' education-only approach. Thereby, we contribute to a better understanding of the effects of personalisation on phishing-related outcomes, such as phishing detection ability, as well as on user perceptions of personalised training.
- Second, our findings underscore the potential of personalisation not only in increasing phishing proficiency to a desired level with enhanced efficiency but also in aligning with users' preferences. Most users stated they prefer personalisation or at least personalised recommendations, with them having the final say in the training selection.
- Third, the chosen personalisation approach was a relatively simple modular method based on phishing proficiency. On the one hand, this indicates that such an approach could be easily implemented in practice with comparably little effort to enhance outcomes and user satisfaction. On the other hand, it serves as a stepping stone towards more sophisticated personalisation to be explored in future research. For instance, emerging AI tools could further enhance personalisation to adjust for industry-specific risks or different job profiles that encounter various threats.
- Finally, we provide the full source code and material used in the training, allowing researchers and practitioners to adopt it themselves.

In the following Section 2 we introduce related work on personalisation in general and related to phishing in particular. Afterwards, we describe the personalisation aspect of the phishing training and the study design in Section 3. Section 4 presents the effects of the personalised phishing training as well as the comparison between the personalised, the randomly assigned and the control condition. Finally, Section 5 discusses the implications of the findings and concludes with recommendations for advancing personalised phishing training.

2 Related Work

In the following, we provide an overview of previous approaches to personalised training and review insights from human-centred anti-phishing training.

Personalisation in Learning Research. In learning research, personalisation has long been identified as a beneficial factor that substantially enhances learning [20], as it accommodates individual differences [14], increases engagement [20], and shows better outcomes than 'one-size-fits-all' approaches [11]. This benefit can be supported with technology [39], showing moderate but positive effects in enhancing education [20, 40], and enhanced interaction with emerging technologies like robots [37]. For instance, Klašnja-Milićević et al. [31] found that a learning software based on learning style and pre-existing knowledge substantially enhanced test scores compared to a non-personalised control. Yet, there is a lack of HCI user studies to clarify the impact of personalisation [79]. Furthermore, different applications of personalisation [27, 68, 77] lead to inconsistencies. For example, personalisation is often used interchangeably with other terms like customisation, adaption, or precision learning [17, 20, 31, 63, 68]. While some approaches use personalisation to refer to reactive difficulty variations based on performance [27, 77], or lower-level adaptions [68], we understand personalisation to first identify the learners' needs and then tailor interventions based on those needs [16]. Therefore, we characterise a personalised system as adjusting to the characteristics of the learner, such as skills and prior knowledge [2, 17, 31], allowing the system to adapt training to optimally challenge and educate the learner [17, 31, 63].

Personalisation in Cybersecurity. Building on the promising results of personalised learning in the educational context, the personalisation of learning materials has also been introduced in cybersecurity. Seda et al. [55] implemented personalisation to improve the learning of cybersecurity topics based on prior knowledge and performance metrics by adapting the difficulty of tasks. They found that personalisation increased participants' training success rates and left students less overwhelmed compared to a control group. However, they only adapted the difficulty of tasks that otherwise targeted all students the same way, rather than accounting for a broader range of characteristics.

The application of personalisation approaches extends to the phishing context. Marforio et al. [41] evaluated the effectiveness of personalised security indicators that users could configure themselves to enhance personal relevance. They found that these personalised indicators significantly reduced phishing susceptibility compared to generic indicators with no personal connection to the users. Roepke et al. [50] used personalisation to customise the links contained in a phishing game to the users' background, thereby increasing their familiarity with the emails they encountered. This approach could increase the realism of phishing tests and training, by tailoring it to the individual learner's usual environment. However, the inclusion of this personalisation did not significantly affect the participants' performance. Zahedi et al. [78] investigated a personalisation approach for security warning indicators against phishing websites, with the aim of enabling users to build trust and familiarity with the indicators. Interaction with the personalised tool enhanced self-reported self-protection behaviour.

While these approaches have focused on personalising messages or content based on performance or other user characteristics, they have not evaluated users' needs or tailored interventions to fit them. Small adaptions, like personalising links, might make interventions more accessible or realistic for specific user groups but appear insufficient to impact performance-related outcomes. Although these preliminary findings are promising, personalisation in cybersecurity has the potential to go beyond simple adjustments like difficulty levels, custom images, or limited variations in training emails. By incorporating users' prior knowledge and skills, we aim to enhance the effectiveness of personalised learning strategies, which forms the basis for the design of this study.

Towards Enhanced Personalisation in Phishing Training. Jampen et al. [29] conducted a comprehensive analysis of factors affecting phishing training effectiveness, highlighting personalisation as a key element due to differences in user capabilities. However, efforts in cybersecurity to provide personalised training are hampered by limited resources [1] and a lack of knowledge on how to personalise. While some studies directly compare different training options (e.g., [13, 56]), they assign these variants randomly. In general,

phishing training still widely uses generic material to educate all users equally [1] and therefore suffers from low engagement [25].

Vasileiou and Furnell [67] highlight the mismatch that results from inter-personal differences, such as knowledge or security awareness. Furthermore, they found that the 'one-size-fits-all' approach is prevalent in most phishing training. Consequently, Alotaibi et al. [2] proposed a personalised security awareness program framework, outlining that user-specific factors, such as prior knowledge or perception of security, can be evaluated in the first step and subsequently used for modular training components. However, the framework does not specify how this personalisation should be implemented. Likewise, previous research did not evaluate the tailoring of phishing interventions based on user information. The present work attempts to address this gap through personalised phishing training.

Very recently, Schöni et al. [54] evaluated the effectiveness of assigning participants to training groups based on phishing proficiency scores. In a study with 96 participants, they found that participants in the lower proficiency training variants showed higher proficiency increases. While their categorisation process was exploratory and led to substantial variations in group sizes, their work nevertheless demonstrated the feasibility of a composite proficiency score for personalisation. The present work aims to build on this exploratory work to reveal personalisation benefits more systematically by including two different control comparison groups, a revised assignment process, and more consistent and validated training material.

Accounting for Users' Skills, Knowledge, and Needs: The Security Learning Curve. To personalise training, it is important to consider the pre-existing skills, knowledge, and needs that different groups might have. Therefore, we designed our personalised phishing approach based on the Security Learning Curve by Hielscher et al. [26], which integrates prior work from [10, 49], to function as a framework for users' needs in cybersecurity learning. As described by Sasse et al. [52] and illustrated in Fig. 2, the model posits a series of steps that build on top of each other, eventually leading to secure behaviour once proficiency has reached a suitable level. For proficiency to increase, different elements are necessary at each step, which likewise benefit from differences in how training is conveyed. As marked in Fig. 2, we focus on three aspects: (a) education to enhance information, sensitisation, understanding and knowledge, (b) practical training to foster skills, ability, and self-efficacy, and finally (c) reminders to support embedding and repetition. For the personalisation of our study, we will therefore employ proficiency-based personalisation that presents content to match the respective proficiency level and needs regarding cybersecurity learning.

3 Method

We conducted a pre-registered, online within-between subjects study, comparing pre- and post-training effects within participants, as well as conditions and proficiency groups between participants. The study was conducted in English. We validated the study concept and procedure in a prior study with 96 participants [54], which led to improvements to the training material and assignment process. The final process involved categorising participants based on a single phishing proficiency score and then assigning them to one of three training groups based on a pre-training questionnaire. We again tested the questionnaire with a pilot of 10 participants to verify the design of the survey and the comprehensibility of the questionnaire. Below we describe our sample, training material, scoring and assignment process, study procedure, and ethical considerations.

3.1 Participants

We initially recruited 346 participants through Prolific, who were proficient in English and located in the United States or Europe. We excluded 4 participants whose completion time exceeded the mean (41 minutes) by more than three standard deviations (95 minutes or longer). All participants passed multiple attention checks and indicated that they participated in good faith. Thus, the final sample comprised N=342 participants. The participants' age distribution was as follows: 60 were between 18-25 (17.5%), 133 between 26-35 (38.8%), 67 between 36-45 (19.6%), 53 between 46-55 (15.5%), 20 between 56-65 (5.8%), 7 between 66-75 years old (2%), 1 above 75 years old (0.3%), and 1 person preferred not to say (0.3%). 181 participants identified as female (52.9%), while 158 identified as male (46.2%), and 3 (0.9%) provided no information on their gender. The participants' highest level of education is as follows: 11 participants had a PhD degree or similar (3.2%), 195 had a graduate university degree (57%), 32 had an associate or technical degree (9.4%), 94 had a secondary school diploma (27.5%), 4 completed primary school (1.2%), and 6 had another type of degree (1.8%). This composition is less representative of the broader population, but is more typical of office workers, who are highly subjected to phishing. We also asked participants about whether their education background was in IT or a related field. 262 participants did not have any such relation (76.5%), 14 participants were IT or IT security specialists (4.1%), 41 participants encountered computer science or IT security during their studies (12%), and 25 participants had another IT security-related education or occupation (7.3%). We additionally asked participants whether they had completed any cybersecurity training before the study. 184 participants never completed a cybersecurity training before (53.7%), 75 completed a cybersecurity training once (21.9%), and 83 completed more than one cybersecurity training (24.2%).

3.2 Training Material

Based on the security learning curve [52], our training consisted of three modular elements: (1) videos as an educational element, (2) an interactive quiz as a practical phishing training element, and (3) nudging banners as a reminder element (see Figure 2). While these elements differed in their method of delivery, they all addressed the same content to enhance comparability across training groups assigned to different modules: All training variants covered mail attachment detection, link detection, and awareness for cognitive biases and heuristics exploited by attackers. These biases and heuristics are mental shortcuts that attackers try to abuse, such as by creating urgency or inducing authority to manipulate users into hasty, non-systematic actions [12]. The learned suspicion from the training is therefore effective at facilitating further critical thinking to investigate links or attachments more systematically [12, 24, 29, 70], mimicking how experts detect phishing [73].

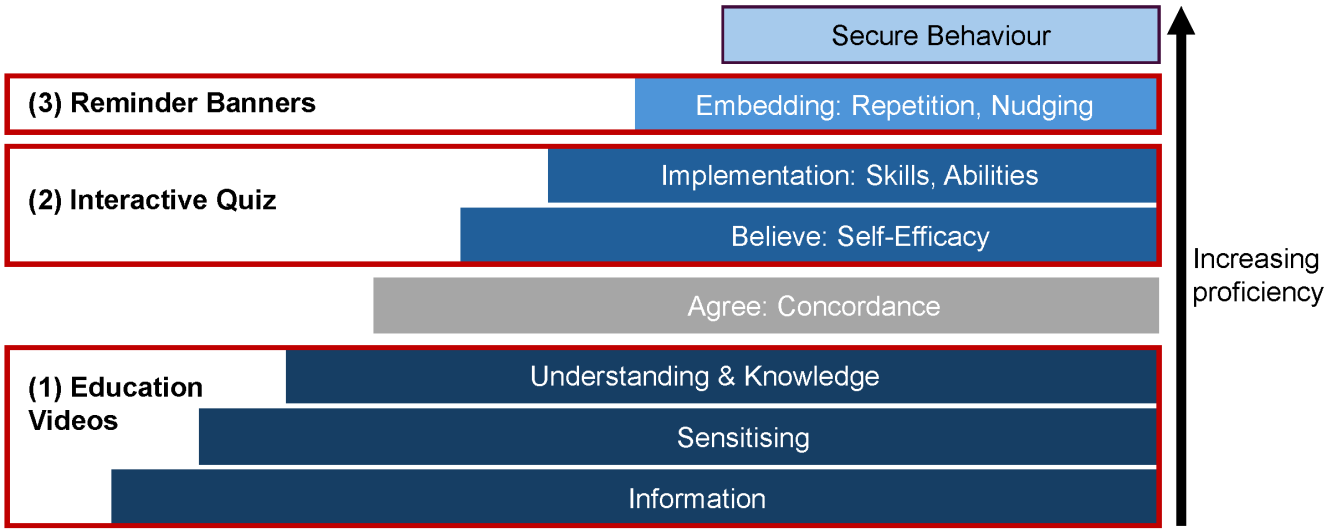


Figure 2: An illustration of the security learning curve adapted from [26] and described in [52]. The model describes a series of steps, building on each other, to reach secure behaviour. The steps addressed in the study, i.e., education, practical training, and reminder, are marked through a red frame.

While validated video material covering that content exists, no pre-existing quiz or reminder material covered all these elements. Therefore, it was necessary to create novel material to ensure consistency between the training variants. We adapted the background task and superimposed banner from previous work [54]. Based on this first implementation, we made further adjustments and incorporated logos that are semantically related to the text content (an example can be seen in Fig. 11 in Appendix A) from a separate study involving 117 participants. The interactive quiz was a novel development, and, together with the other finalised training elements, was evaluated in an internal pilot study involving 5 experts and 7 lay users, as well as in a subsequent pilot study with 10 participants. We show examples of these elements in Fig. 3 and provide the source code for the modular training and its materials on GitHub.[1] In the following, we briefly explain these elements.

The three **educational videos** we used are part of the NoPhish series, developed to educate users about phishing in an engaging way [72]. They have been selected because of their scientific foundation, their coverage of prominent phishing topics and their successful empirical evaluation in previous work (e.g., [7, 8]).

In the **interactive quiz**, participants processed a series of example emails and decided between three options which action seemed most prudent (such as reporting an email, deleting it, or replying to it). This decision could be based on features in the emails, such as their framing, links, attachments, or sender information. After each choice, participants received direct feedback, highlighting elements in the email that might confirm or contradict their decision. This exercise was similar to other practical training that develops and enhances detection skills [22, 51].

[1]GitHub Repository: https://github.com/lorinschoeni/personalised-phishing

The **nudging banners** incorporated a combination of icons and text designed to remind users of various techniques for identifying phishing, thereby helping to maintain a heightened level of alertness. This approach aligns with training methods that focus users' attention on potential threats by providing targeted warnings through in-situ interruptions of normal workflows [22, 51]. They were designed to be displayed at specific intervals while participants were engaged in a background task. Rather than relying on direct examples, such periodic reminders of the general threat can effectively sustain high levels of phishing awareness [34, 48].

3.3 Training Assignment & Personalisation

We assigned all participants to one of three conditions at random, which in turn defined how participants were assigned to specific training variants. In the personalised condition, participants were assigned based on their proficiency, as described in Section 3.3.1. Additionally, we assigned participants to a randomised or education-only control condition.

In the randomised control condition, participants were randomly assigned to one of the training variants irrespective of their proficiency score. For example, a participant with a high proficiency score might be randomly assigned to the low proficiency training and therefore complete all elements, including the educational videos, the interactive training and the banners during the background task. The selected control condition enabled us to evaluate whether the personalised assignment achieved improved outcomes in comparison to a random assignment, which does not take into account the participants' existing knowledge and skills. In the education-only control condition, participants watched the educational videos but did not engage in any other training component, including the background task. This allowed us to compare our results to simple training, which functions as a baseline. The following summarises differences between the conditions:

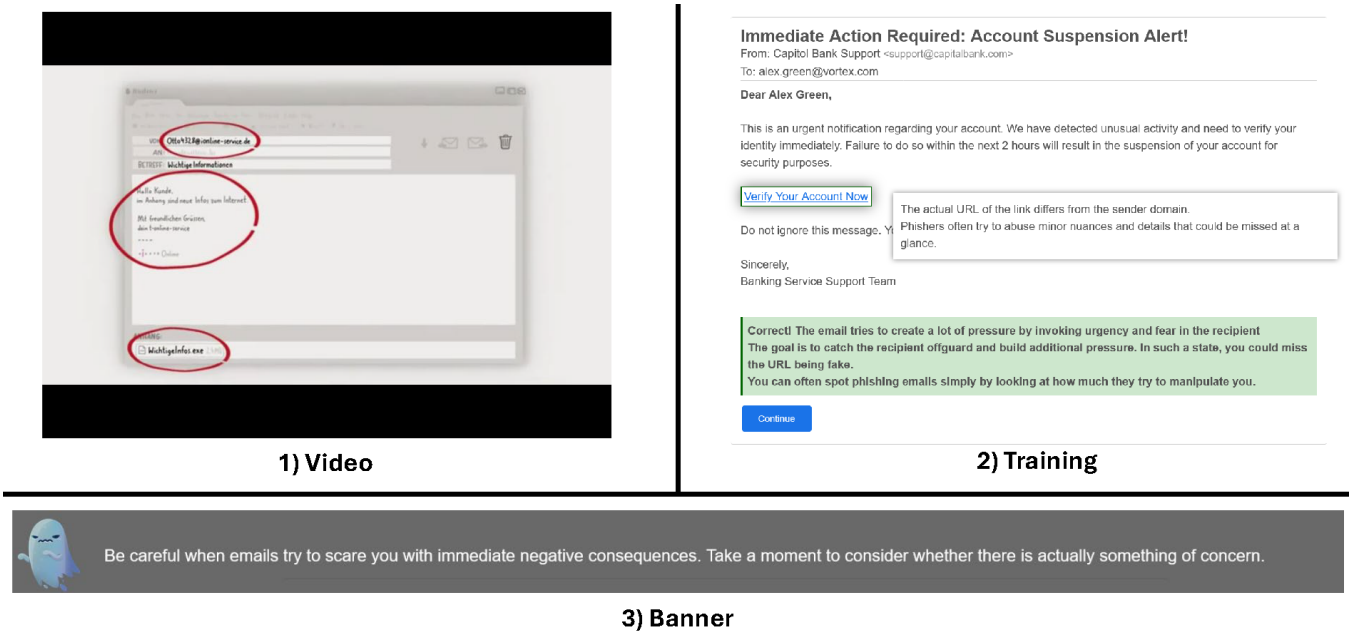


Figure 3: Overview of the training elements that included 1) a screenshot from one of the educational NoPhish videos shown to participants, 2) an exemplary email from the interactive phishing training teaching users how to detect phishing, and 3) an example of a banner used to remind participants during the background task in which participants used an emulated email client.

- **Personalised Training** (experimental condition): Participants in this condition were assigned to the appropriate training variant based on their proficiency level.
- **Random Assignment Training** (control condition): Regardless of their proficiency scores, participants in this condition were randomly assigned to one of the three training variants.
- **Education-Only Training** (baseline control condition): Participants in this condition only watched the educational videos.

In Table 1, we provide an overview of participant assignments to each condition and training variant. As proficiency followed a normal distribution, we assigned a higher percentage into the personalised condition to ensure a large enough sample size in the low and high proficiency training variants.

Table 1: Summary of how many participants were assigned to each training variant. Participants in the personalised condition were assigned based on pre-training classification.

Group	Personalised	Randomised Control	Education Control
Education-only	-	-	89
Low	37	28	-
Medium	75	31	-
High	46	36	-

3.3.1 Personalisation. In the personalised training condition, we categorised participants into three groups based on their pre-training proficiency, similar to previous studies (e.g., [33]) and in alignment with proposals for NIST's NICE framework [61]. This approach enabled us to capture multiple dimensions of phishing-related proficiency—such as knowledge, ability tests, self-assessments, and attitudes—while using a score as an intuitive measure of overall proficiency and progress along the Security Curve.

To account for the multifaceted nature of proficiency, we employed several scales that assess relevant metrics and differentiate between individuals [42], adapted from [54]. Given the lack of scales targeting phishing-specific expertise, we utilised the established SA-13 scale [19] to measure general cybersecurity attitude. To capture knowledge, we employed email use and knowledge items from the Human Aspects of Information Security Questionnaire's (HAIS-Q) [43], which has been constructed to be modularly used. However, as the HAIS-Q does not directly address phishing, we complemented it with phishing-specific knowledge tests. We additionally collected self-estimates of knowledge, ability, and alertness to capture participants' self-image of where they stand on the Security Curve, as well as a phishing classification task to directly evaluate participant abilities. We further elaborate on how these variables are measured in Section 3.4.

For the personalisation, we further added information on email use and previous cybersecurity training experience to better understand participants' backgrounds. These scores were all summed up into a single **participant score**. We provide an overview of

this score calculation in Fig. 4. Based on the resulting score, participants were categorised into one of three different groups. If their score was below 20, they were assigned into the low proficiency group. If their score was above 26, they were assigned into the high proficiency group. Everyone else was assigned into the medium proficiency group. The cut-offs were based on insights from [54], with adjustments after internal testing and a pilot test with 10 participants. The final sample's proficiency score distribution (Median = 23.1, SD = 4.46) indicated that our cut-offs (proficiency scores of 20 and 26, respectively) were suitable for capturing medium proficiency and separating it from low and high proficiency groups. We summarise group differences in the following:

1) **Low proficiency:** In line with the security learning curve (cf. Fig. 2) people with low proficiency first need to learn, e.g., through ***educational video material***, what phishing is, why it is relevant, and how to recognise it. Then, the transfer of the theoretical knowledge to everyday life situations, i.e., the implementation of skills and abilities [52], needs to be practised. Therefore, the group additionally completed an ***interactive phishing detection quiz***. And finally, in everyday life the alertness for phishing might decrease over time or in stressful situations. Repetition and nudges such as reminders can support successful habituation of the learned behaviour [52]. Hence, the group finally received ***reminders***, informed by the findings of [48] to keep alertness levels high.
2) **Medium proficiency:** People with medium proficiency know about phishing in theory but may lack practise. Hence, they only completed the ***interactive phishing detection quiz*** and also received ***reminders*** to keep alertness levels high.
3) **High proficiency:** People with high proficiency know about phishing and can successfully detect phishing emails. However, even they might lack alertness in everyday life, preventing them from successfully applying their skills. Hence, this group only received ***reminders*** to counteract a lack of alertness.

In total, there were 82 participants with low proficiency, 157 participants with medium proficiency, and 103 participants with high proficiency. As proficiency follows a normal distribution, a higher proportion exhibited medium proficiency. In Table 2, we provide a more detailed overview of proficiency groups across the different conditions. Only participants in the personalised condition were assigned to training based on their proficiency level.

3.4 Procedure

After providing informed consent, participants first completed a pre-training questionnaire, then the training they were assigned to, and finally a post-training questionnaire. The entire study, including the training, was hosted on Qualtrics. The pre- and post-training questionnaire both contained sections to test specific components of phishing proficiency, including self-estimates of phishing proficiency, knowledge tests, the Security Attitude Inventory with 13 items (SA-13) [19], and a phishing classification task including screenshots of emails that needed to be classified as legitimate or phishing emails. In addition, in the pre-training questionnaire, we collected phishing-related background from knowledge items in the email use and knowledge focus areas of the HAIS-Q [43]. In the post-training questionnaire, we collected training feedback, all items from the HAIS-Q's email and internet use focus areas, as well as demographic data, including personality aspects using the short version of the Big Five Inventory with ten items (BFI-10) [47]. The full questionnaire can be found in the supplementary material. The phishing knowledge test consisted of 5 multiple-choice questions in both tests, evaluating participants' knowledge, where one out of multiple answers was correct. One example of a question is "If you receive an email that appears to be from your bank, asking for your password, what should you do?" In the phishing classification task, we presented 12 emails to participants in both tasks, where participants were prompted to classify whether each email is phishing or not. These emails were interactive, i.e., users could learn more information about the sender or hover over links. The order of all phishing knowledge and classification test questions was randomised across the study, to account for potential differences in difficulty. The procedure is visually summarised in Fig. 5.

Training. After the categorisation, participants interacted with the training elements based on their condition and assignment. To increase realism and keep all participants occupied for the same amount of time despite different training components of varying length, participants in the personalised and randomised conditions completed a background task during the training. This task involved interacting with emails in a fictional office-themed mailbox with various actions, such as downloading or opening an attachment, reporting an email, or sending a reply. We provide more information on the background task in Appendix A. The educational video and interactive training elements temporarily paused the background task, ensuring that participants focused exclusively on these components until they were completed. Once finished, participants resumed the background task, continuing until another training element was triggered or the training ended. In contrast, the reminder banners were integrated into the background task and displayed at set intervals, allowing participants to encounter all banners during the training. After the training, participants were asked for feedback on the training, followed by the post-training questionnaire.

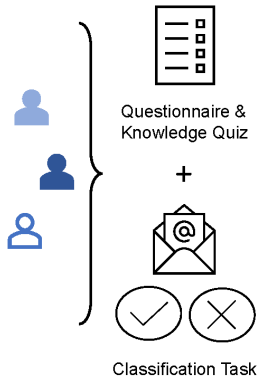
3.5 Ethical Considerations

The design of our study followed established ethical standards for psychological research involving humans [3] and was approved by our university's IRB. We minimised the potential for privacy invasion, e.g., by collecting age ranges instead of a concrete age. Participants were informed about the nature of the tasks and provided informed consent. Participation was voluntary and participants could abort the study and request the deletion of their data at any time without negative consequences. All participants received an equal payment in line with Prolific's suggestions for fair compensation of £6.75 for their 45-minute participation.

3.6 Pre-Registration

This research has been pre-registered on OSF to enhance transparency and replicability of the work. [2]

[2]Pre-registration: https://osf.io/ub5jp



Component	Amount	Explanation	Score 1	Score 2	Score 3
Email client use*	1 question	1 point for "Multiple times a day", 2 points for "Multiple times an hour"	1	1	2
Training Experience*	1 question	2 points for "Yes, once", 4 points for "Yes, more than once"	0	2	4
Self-reported proficiencies	3 dimensions	2 point for scale point 5, 1.5 points for scale point 4, and 1 point for scale point 3	3	3.5	5.5
HAIS-Q**	6 Items	0.5 points for scale point 5, 0.2 points for scale point 4, and 0.1 points for scale point 3	1.3	2	3
SA-13**	13 Items	0.5 points for scale point 5, 0.2 points for scale point 4, and 0.1 points for scale point 3	1.8	3.1	4.5
Phishing Quiz	5 Questions	1 point for each correct answer	2	3	4
Email Classification	12 Emails	1 point for each correct answer	7	9	11
Total Score			**16.1**	**23.6**	**34**

* These components have been used for training assignment (pre-training) but not for the proficience score calculation reported in the analysis, as they are unaffected by the training.
** Negative scale items have been inverted prior to score calculation.

Figure 4: An overview of the training score assignment based on [54]. All participants completed the questionnaire, knowledge quiz and classification task. We calculated a score based on their email client use, training experience, self-reported proficiencies, their security awareness, their security attitudes, their correct answers in the phishing quiz and correct answers in the email classification task. The table shows detailed explanations of the score components and attribution with example calculations.

Table 2: Summary of how many participants were assigned to which condition and their distribution of proficiency levels as calculated from the pre-training score based on a questionnaire and classification task.

Group	Personalised	Randomised	Education-Only
Total Number	158	95	89
Distribution of proficiency levels based on pre-training questionnaire score			
Low	37	24	23
Medium	75	46	38
High	46	25	28

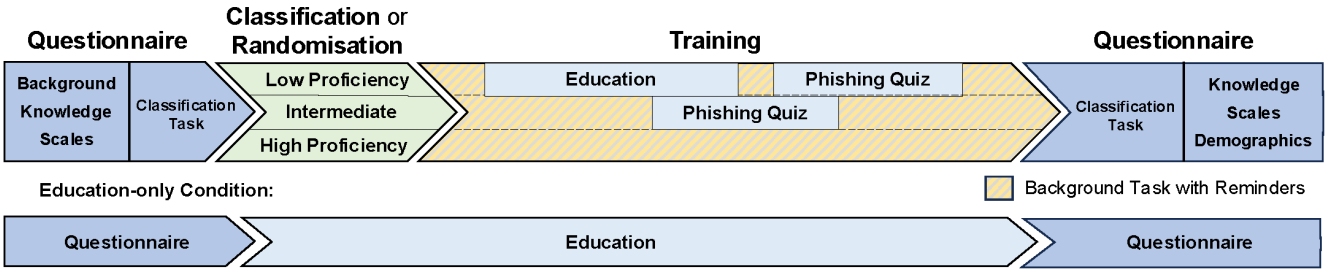


Figure 5: An overview of the study procedure and the three conditions. All participants first went through a questionnaire and classification task, and subsequent clustering into three proficiency levels: low, medium, and high. Participants in the education-only control condition only watched the educational video material. After the training, all participants again completed a questionnaire and classification task to measure their post-training proficiency level.

4 Results

In the following, we first present the results in line with our pre-registered research questions and hypotheses. Afterwards, we report on additional exploratory analyses.

All quantitative data analysis was conducted in R version 4.3.2 [46]. We calculated regression models using the *lme4* package [6]. Specifically, we fit a linear model to assess the effects of time (pre- vs. post-training), assignment condition, and proficiency level on the

variables of interest as outcome variables. We then conducted post-hoc tests using the *emmeans* package [36] to determine the estimated marginal means of the interaction between time (pre- vs post-training), proficiency level, and training assignment condition. We contrasted these marginal means to test hypotheses H1 to H4. This allowed us to isolate and compare the effects of each factor while controlling for the influence of other variables in the model. We used an adjustment to account for the increased probability of false positives in multiple testing, which can increase the chance of false negatives. In the following sections, we report only the relevant contrasts derived from the post-hoc analyses. The full results of the regression models, including detailed parameter estimates, can be found in Appendix B.

We applied an alpha level of 0.05 as the significance threshold. In our analysis, we used confidence intervals to provide more nuanced insights into the range of potential effects. A confidence interval that includes zero indicates a non-significant p-value. We did not transform the different variables, which ensures that they represent original scales and observed behaviour. While this limits direct comparability between variables, the results are therefore directly interpretable. For instance, a 0.05 difference in phishing detection accuracy represents a 5% difference in the accuracy of detecting phishing emails or a 0.1 difference in security attitude represents a 0.1 difference in the SA-13 scale's score. Consequently, we group the reporting of results by the different variables.

We also assessed whether the control variables age, education, or one of the BFI personality traits significantly influenced proficiency gain. To do so, we fitted competing models with each control variable as a fixed factor. However, none of these models improved the fit (see Appendix B). Consequently, we excluded the control variables from further analysis.

4.1 Training Increases Proficiency Across Groups

Overall, the training increased the phishing proficiency of all groups. These effects extended to the components making up the composite proficiency score, including phishing detection accuracy, security attitude, and proficiency self-estimates. The only exception is the accuracy of detecting benign emails as benign, which slightly decreased through the training. In other words, participants were slightly more likely to classify a benign email as phishing after the training. We provide an overview of the pre- and post-training effect across all conditions and groups in Table 3.

In the following, we will describe the training effects on proficiency increases and its key component factors, as well as differences between conditions and groups to answer H1 to H4.

4.1.1 Effects on Phishing Proficiency. We analysed the effect of the training on the phishing proficiency score we calculated. Our results support H1 to H3 regarding phishing proficiency improvements. Specifically, participants demonstrated a significant increase in phishing proficiency following the intervention ($M = 2.73, SE = 0.28$, 95% CI [2.17, 3.28]), confirming H1. Furthermore, participants receiving the full training outperformed those in the education-only group, showing significantly greater improvement ($M = 1.95, SE = 0.58$, 95% CI [0.60, 3.29]), in line with H2. Similarly, participants with a lower baseline proficiency improved substantially more compared to those with a higher baseline proficiency ($M = 3.94, SE = 0.75$, 95% CI [2.20, 5.67]), supporting H3. However, H4 was not supported, as there was no significant difference between the personalised and randomised assignment groups ($M = 0.75, SE = 0.46$, 95% CI [-0.32, 1.82]). While the personalised group's estimate was higher, its confidence interval included zero. We summarise these results in Table 4.

4.1.2 Effects on Phishing Detection Ability. We analysed participants' ability to detect emails, which ranges from 0, indicating no correct detection to 1, indicating completely correct detection. Our results for detection ability also support H1-H3. Participants in the personalised condition showed a significant increase in detection ability post-training ($M = 0.19, SE = 0.02$, 95% CI [0.14, 0.23]), consistent with H1. Additionally, participants who underwent full training exhibited significantly greater improvements compared to those in the education-only group ($M = 0.13, SE = 0.05$, 95% CI [0.02, 0.23]), supporting H2. As predicted in H3, participants with lower baseline detection ability improved more than those with higher baseline detection ability ($M = 0.18, SE = 0.06$, 95% CI [0.04, 0.32]).

We further analysed the benign email detection accuracy of participants, that is, how accurately participants classified benign emails as not-phishing. This allowed us to isolate whether the higher phishing detection behaviour actually indicates increases in phishing detection ability rather than general increases in suspicion and thus over-reporting. As shown in Table 6, none of these effects were significant, indicating that the benign email reporting behaviour did not significantly change for the compared groups.

In contrast, H4 was not supported for detection ability, as the difference between personalisation and random assignment was not statistically significant ($M = 0.04, SE = 0.05$, 95% CI [-0.04, 0.13]). We visualise these changes in Fig. 6. While participants in the personalised condition demonstrated a higher increase in the phishing detection accuracy and a lower decrease in the benign detection accuracy, the confidence intervals included zero.

We further break down these changes into proficiency groups for the personalised and randomised training assignment conditions. Fig. 7 shows the training effect on the personalised condition, indicating that benign detection accuracy remains relatively constant, while phishing detection accuracy increases to very similar levels for all proficiency groups. For the randomised assignment condition shown in Fig. 8, we generally observe a higher variance and a tendency for lower post-training benign detection accuracy. Furthermore, the upper and lower boundaries of the confidence interval differ more substantially and again show higher variance, demonstrating a less consistent trend in post-training values.

4.1.3 Effects on Security Attitude and Self-Estimated Proficiency. Our findings on security attitudes, measured through the SA-13, support H1, but fail to support H2-H4. Specifically, participants in the personalised training condition exhibited a significant improvement in security attitude post-training ($M = 0.24, SE = 0.06$, 95% CI [0.12, 0.36]), confirming H1. However, H2 was not supported, as the difference between full training and education-only groups was not statistically significant ($M = 0.14, SE = 0.13$, 95% CI [-0.16, 0.44], p = 0.44). Similarly, H3 was not supported, with no significant difference in security attitude improvements between participants

Table 3: Marginal mean estimates from the pre- and post-training contrasts for the variables from each regression model. Significant values, i.e., for which the confidence intervals (CIs) do not include 0, are marked with an asterisk *.

Variable	Estimate	*SE*	Lower CI	Higher CI
Total Proficiency Score*	2.3	.20	1.92	2.68
Phishing Detection Accuracy*	.15	.02	.11	.18
Benign Detection Accuracy*	-.05	.02	-.08	-.02
Security Attitude*	.27	.04	.19	.36
Self-Estimate Knowledge*	.65	.06	.54	.77
Self-Estimate Ability*	.45	.06	.34	.56
Self-Estimate Alertness*	.66	.06	.54	.78

Table 4: Marginal mean estimates from the regression model including the proficiency level as a variable. These marginal means are contrasted between different participant groups to highlight differences for the four hypotheses H1-H4, with the value denoting the estimated mean difference between the compared groups. Significant values, i.e., for which the confidence intervals (CIs) do not include 0, are marked with an asterisk *.

Variable "Total Proficiency Score"	Estimate	*SE*	Lower CI	Higher CI
H1: Personalisation Improvement*	2.73	.28	2.17	3.28
H2: Full vs. Education-only*	1.95	.58	.60	3.29
H3: Low vs. High Baseline*	3.94	.75	2.20	5.67
H4: Personalisation vs. Randomisation	.75	.46	-.32	1.82

Table 5: Marginal mean estimates from the phishing detection accuracy regression model. The contrasts highlight group differences for the four hypotheses H1-H4. Significant values, i.e., for which the confidence intervals (CIs) do not include 0, are marked with an asterisk *.

Variable "Phishing Detection Accuracy"	Estimate	*SE*	Lower CI	Higher CI
H1: Personalisation Improvement*	.19	.02	.14	.23
H2: Full vs. Education-only*	.13	.05	.02	.23
H3: Low vs. High Baseline*	.18	.06	.04	.32
H4: Personalisation vs. Randomisation	.04	.05	-.04	.13

Table 6: An overview of marginal mean estimates from the benign detection regression model. The contrasts highlight group differences for the four hypotheses H1-H4. Significant values, i.e., for which the confidence intervals (CIs) do not include 0, are marked with an asterisk *.

Variable "Benign Detection Accuracy"	Estimate	*SE*	Lower CI	Higher CI
H1: Personalisation Improvement	<.01	.02	-.04	.05
H2: Full vs. Education-only	.05	.05	-.06	.16
H3: Low vs. High Baseline	.03	.06	-.12	.17
H4: Personalisation vs. Randomisation	.08	.04	-.01	.17

with lower and higher baseline attitudes ($M = 0.36, SE = 0.17$, 95% CI [-0.02, 0.75], p = 0.75). Finally, H4 was also unsupported, as no significant difference was observed between personalised and random assignment groups ($M = -0.06, SE = 0.10$, 95% CI [-0.30, 0.18], p = 0.18).

Finally, we analysed the effect of training improvement on the three phishing self-estimates questions, where participants rated on a scale from 1 to 5 how high they estimated their own ability to detect phishing, alertness to phishing, and knowledge about phishing. In Fig. 9, we provide an overview of the estimates and confidence intervals. Overall, H1 and H3 are again supported for all three variables, with alertness and knowledge showing higher improvements than ability self-estimates. H2 was partially supported for alertness and knowledge, but not for the ability self-estimate, where the difference between full training and education-only groups is not significant. The results do not support H4, as none of the self-estimates differed significantly between the personalisation and randomised assignment conditions.

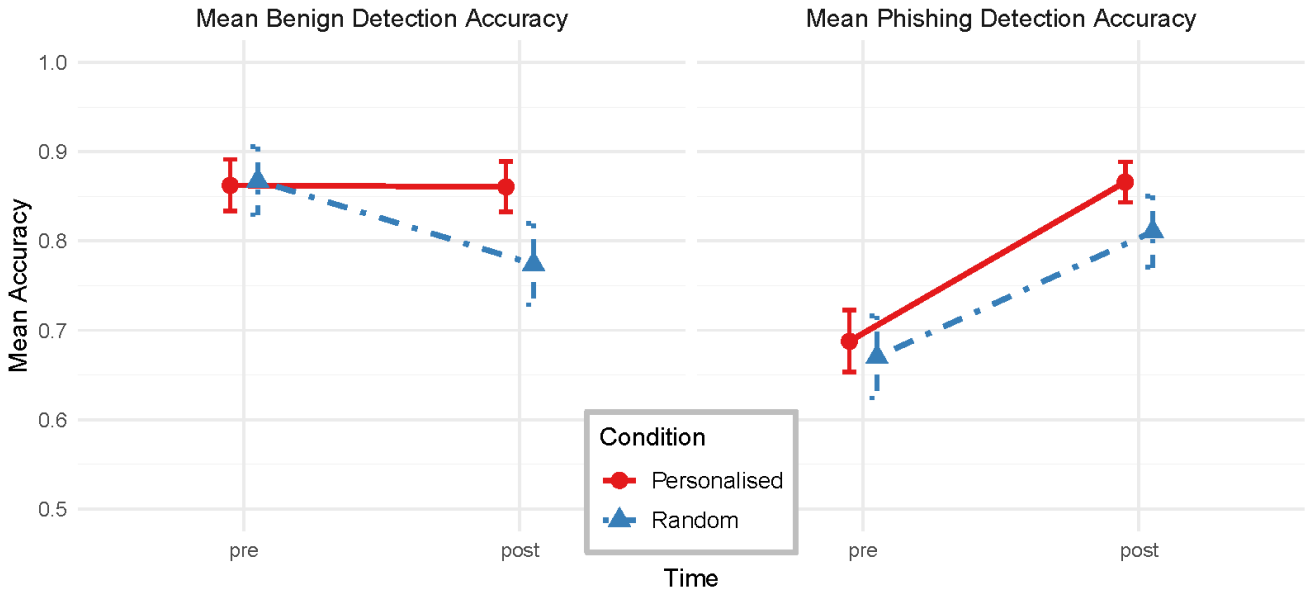


Figure 6: Phishing and benign email detection performance of the personalised as compared to the randomised training assignment conditions.

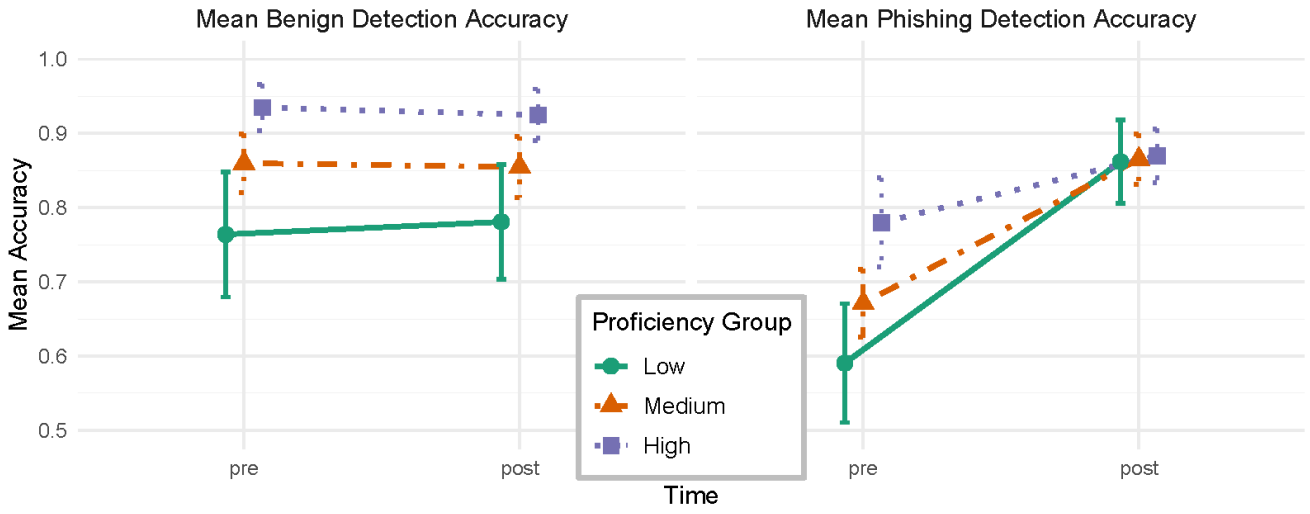


Figure 7: Phishing and benign email detection performance of proficiency groups in the personalised training condition.

4.1.4 Exploratory Analysis of Training Match. Beyond the pre-registered analysis, we conducted an exploratory analysis to evaluate whether user proficiency matching the training variant they received influenced any training outcomes. Many participants in the randomised control condition were assigned to a training variant that corresponds to their proficiency level by chance. Thus, a certain percentage of participants were also correctly matched with training content, which might have influenced the comparison with the personalised condition. To correct for that influence, we exploratively conducted a direct evaluation of the training "match" vs. "non-match". To this end, we calculated a new model with training match as an additional explanatory variable. We calculated this model on a subset of the sample that excluded participants in the education-only control condition, as they all received a baseline training. We then again contrasted the marginal means between groups that either matched or did not match the training variant, which we present in the following. A summary of these results can be seen in Table 8.

For overall phishing proficiency there was a significant difference between the "matched" and "non-matched" participants

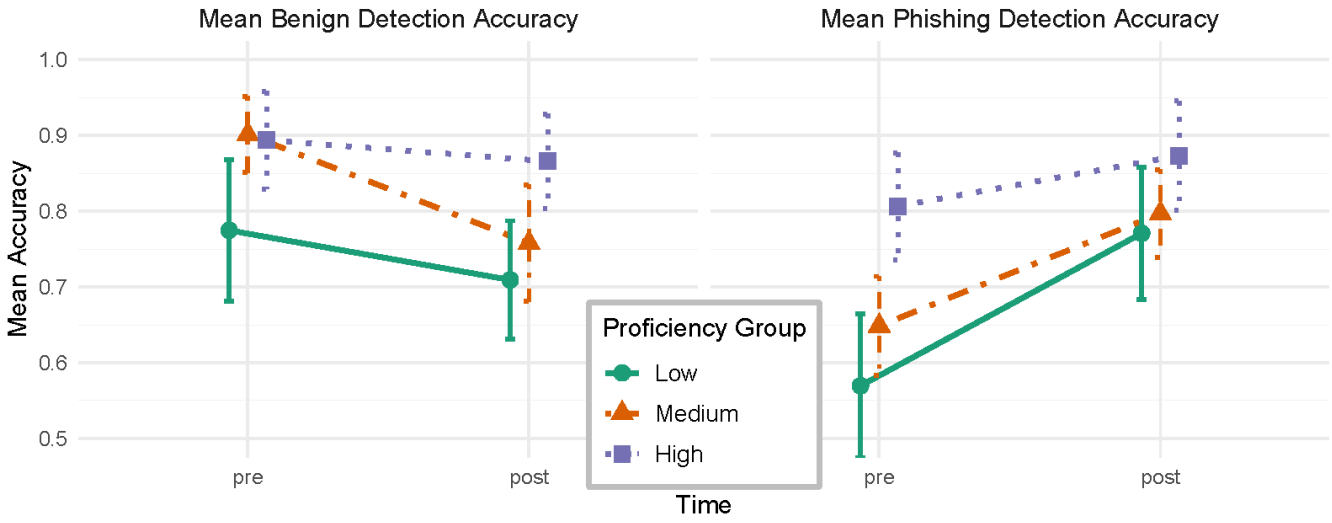


Figure 8: Phishing and benign email detection performance of proficiency groups in the randomised training condition.

Table 7: An overview of marginal mean estimates from the security attitude regression model. The contrasts highlight group differences for the four hypotheses H1-H4. Significant values, i.e., for which the confidence intervals (CIs) do not include 0, are marked with an asterisk *.

Variable "Security Attitude"	Estimate	Standard Error	Lower CI	Higher CI
H1: Personalisation Improvement*	.24	.06	.12	.36
H2: Full vs. Education-only	.14	.13	-.16	.44
H3: Low vs. High Baseline	.36	.17	-.02	.75
H4: Personalisation vs. Randomisation	-.06	.10	-.30	.18

($M = 1.03, SE = 0.34$, 95% CI [0.36, 1.7]). Furthermore, we found significant differences in phishing detection accuracy ($M = 0.06, SE = 0.03$, 95% CI [0.01, 0.16]), self-estimated knowledge ($M = 0.23, SE = 0.10$, 95% CI [0.03, 0.43]), self-estimated ability ($M = 0.20, SE = 0.10$, 95% CI [0.001, 0.40]), and security attitude ($M = 0.19, SE = 0.07$, 95% CI [0.04, 0.34]). However, we found no significant difference for benign detection accuracy and self-estimated alertness.

4.2 Preference of Personalisation

Type of Training Assignment. After the training, we asked participants how they preferred to be assigned to different phishing training variants. We asked them if they prefer automatic assignment based on personalisation, a recommendation with final choice, or simply self-selection with no additional input.

Consistent with H5, participants favoured personalisation over recommendation. Generally, participants favoured personalisation as a recommendation, while keeping the final choice. A sizeable amount of participants also preferred automatic personalisation with no user input. We evaluated whether this preference differed based on whether the training variant they experienced fits their proficiency group, with an overview provided in Fig. 10. A Pearson's Chi-squared test confirmed that this feedback was significantly affected by this match ($\chi^2(6) = 17.78, p = .006$). If people were put into a lower proficiency training than their own proficiency group would have determined or if it was matched, only a small number of participants preferred to select the training themselves. However, if people experienced a training that was tailored towards a higher proficiency group, a large number of participants instead preferred to select the training themselves.

Training Variant Preference. We additionally asked participants whether they preferred the training they just completed, compared to all other alternatives they could have experienced. We evaluated whether this preference depended on the training variant's match with the participants' proficiency group. A Pearson's Chi-squared test confirmed that this feedback was significantly affected by whether the training matched ($\chi^2(9) = 29.27, p < .001$). These results support H5a, as people preferred the training they experienced compared to other options. Accordingly, people preferred the training they experienced more if they were assigned through personalisation rather than randomly.

4.3 Training Design

Immediately after the training, we asked participants for their feedback in three open-ended questions: What they liked about the training, what aspects they did not like, and what else they would

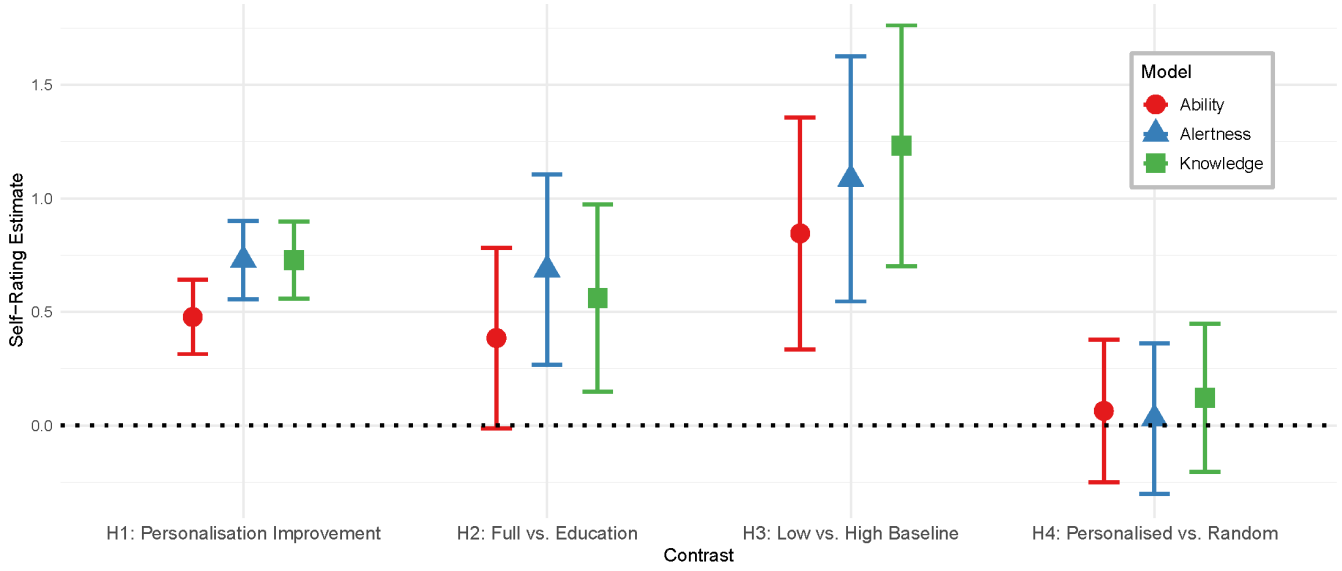


Figure 9: Overview of the self-rating contrast estimates for each hypothesis, across all three variables: ability, alertness, and knowledge. Ratings were given on a scale from 1 to 5.

Table 8: An overview of estimated marginal mean differences across all regression models. The contrasts highlight differences between participants that were assigned a training variant "matching" compared to "not matching" their proficiency level. Significant values, i.e., for which the confidence intervals (CIs) do not include 0, are marked with an asterisk *.

Variable	Estimate	*SE*	Lower CI	Higher CI
Total Proficiency Score*	1.03	.34	.36	1.7
Phishing Detection Accuracy*	.06	.03	.01	.16
Benign Detection Accuracy	.02	.03	-.03	.07
Self-Estimated Knowledge*	.23	.10	.03	.43
Self-Estimated Ability*	.20	.10	.001	.40
Self-Estimated Alertness	.19	.11	-.03	.40
Security Attitude*	.19	.07	.04	.34

like to see in a phishing training. These answers were inductively coded by three independent raters. After completing 10% and 100% of the coding each, the raters compared their coding and resolved any disagreements by discussion. In the following, we summarise these responses. We indicate participants by their ID and a starting character to describe their proficiency level, with 'L' indicating low, 'M' indicating medium, and 'H' indicating high proficiency.

4.3.1 General Positive and Negative Feedback. **"Fun to interact with (M137)"** Participants across all training groups enjoyed the interactive nature and realistic setting of the training, and found it to be informative, interesting, and engaging. The clear and concise explanations and clear terminology, including in the videos, made it easy for participants of all levels to understand. They found that the new information had an impact on their awareness. For instance, one participant mentioned "it gave me useful and relevant information about phishing, a lot of which I wasn't previously aware of. (L147)"

Repetition and Pace. Participants disliked the repetition of some elements, such as the banners, the completion of the background task, or the videos if they were already familiar with them. While participants predominantly reported the pace as being appropriate and the information as being easy to understand, some felt that the pace of the training was too slow and that the training as a whole took too long to complete. For instance, a participant in the education-only condition described that "the training assumes you don't know much about Phishing and talks to all levels of viewers, but this was tiresome for someone who knows more about this subject. (H9)" Another participant in the full training mentioned "Videos were too long and could have got the point across more quickly. Potentially because I'm quite aware of phishing scams, so it covered ground I was already familiar with. (M107)"

4.3.2 Feedback on the Training Material. While the main focus was on investigating the benefits of training personalisation, we also evaluated participant perception of the training components we used in this context.

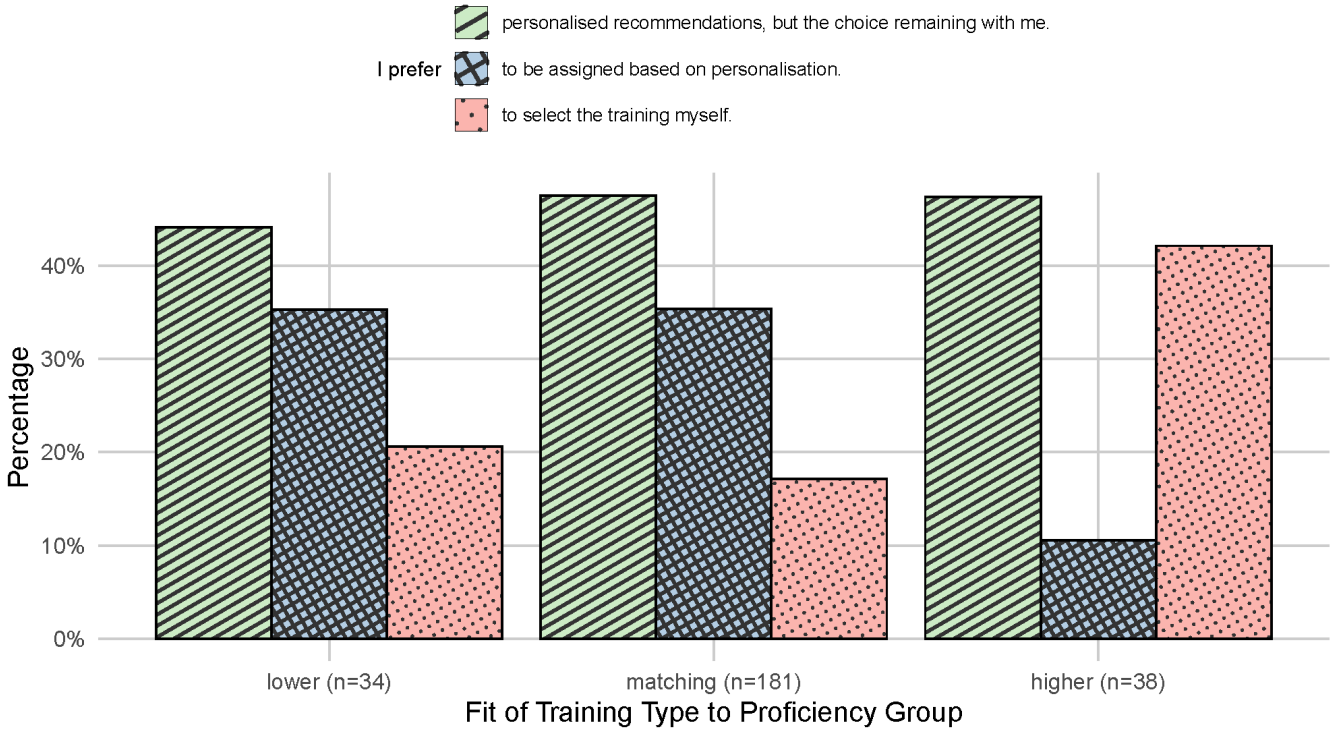


Figure 10: Participant feedback on how they wish to be assigned to personalised training. We show whether the assignment matched their proficiency level (matching), was below their proficiency level (lower), or above it (higher).

Educational Videos The NoPhish videos were overall well perceived by the participants. They enjoyed the animation style and perceived the narrator as pleasant and at eye-level, able to clearly and concisely present technical information. As one participant mentioned, "The videos were engaging and entertaining, without trivialising the topic (H277)." The integration of stories was particularly well-received by participants, with one expressing "I liked the storytelling aspect of the videos. They made the content interesting, compelling and watchable — I didn't struggle to keep focused on the content. (H201)" Nevertheless, some participants criticised that the videos were too slow and long, with several suggesting that they should be consolidated. Additionally, the use of the German context was described as unfitting in an international setting.

Interactive training The quiz-like interactive training, where participants made choices and received feedback on their answers, was particularly well received. Many participants mentioned that it helped reinforce prior knowledge learned in the video, with participants describing that "the interactive choices were fun and drove home the message. (L252)" and that they "allowed me to apply the new cybersecurity knowledge that I learned from the videos. (L26)"

Banner Reminders Participants also mentioned that the reminders during the background task also kept their awareness high throughout the training. Some participants mentioned that they "felt non-intrusive yet provided me with some timely reminders about my behaviour. (H41)" The banners were sometimes described by participants as annoying or distracting. In particular, a few participants mentioned that they would rather see direct examples of emails, rather than banners of the threat in general. One participant described "I didn't like the pop-ups at the top. I would prefer to see examples of these emails. (M309)"

4.3.3 Recommendations. **Adapt Training.** Participants prominently mentioned that they would have liked to see additional videos especially from other perspectives, such as "from victims of phishing and leading experts advice. (M175)" Furthermore, participants indicated a preference for being taught about other threats and scenarios, like smishing or romance scams. For instance, one participant wrote that they wanted "more scenarios, to further explain the dangers and what to look out for. (M238)"

The training's interactivity was well perceived and many participants stated a preference for additional knowledge quizzes or detection tasks. Various participants would have liked either immediate feedback on performance or an overall report at the end of the training. They felt this would give them an additional edge by clarifying what is and is not phishing through additional practice. For instance, one participant mentioned that they wanted "a quiz at the end to check on learning. (H133)"

General Differences Between Proficiency Levels. Participants generally requested more information on what elements would make emails suspicious and what appropriate action could be taken when they encounter a phishing email. Lower proficiency

levels wanted additional information on how they could prevent phishing, while higher proficiency levels desired information on how to recover after falling for phishing. Participants of higher proficiency levels also requested more technical details such as information about malicious file extensions or on suspicious message headers.

4.4 Summary of Results

In the following, we summarise the results of the quantitative and qualitative analysis. Across the hypotheses, we found that personalised training generally led to improvements in most areas, especially for participants with lower baseline proficiency. Benign email detection accuracy was a consistent exception, showing little improvement across all conditions but also no decrease. Personalisation was preferred by participants, and while there was no significant performance difference between the personalised and random assignment condition, there was a significant difference on whether training matched or did not match users' proficiency levels. The full training was generally more effective than education-only approaches, except in a few areas like benign email detection and self-assessed phishing detection ability. Overall, personalisation and proficiency-matched training were favoured and showed greater impact on improvement. The findings relating to the hypotheses are summarised in Table 9.

Participants provided generally positive feedback after the training, consistently across proficiency levels but more critical in the randomised condition. Participants appreciated the interactive and realistic training, finding it informative and easy to understand. The videos and interactive quizzes were generally well-perceived and especially useful in combination, while the reaction to the banners was still positive, but more mixed. However, participants also reported issues with the pacing and scope, occasionally finding the training too slow or lengthy. Participants recommend incorporating more expert videos and covering more types of phishing, along with adding interactive quizzes and providing immediate feedback. Additionally, they suggest including more in-depth information on phishing consequences, tailored to different proficiency levels.

5 Discussion

In the following, we discuss the training effectiveness, the benefits of personalisation, and the training material. We conclude by providing recommendations for phishing training based on our results, followed by a reflection on limitations and potential for future work.

5.1 Effectiveness of the Training

Overall, the training was seen as effective and all groups across conditions displayed substantial improvements to their phishing proficiency. Confirming H1, the personalised groups demonstrated potential gains in the self-perception of their phishing skills, their security attitude, as well as their behaviour.

Control variables show no significant effect. We measured several control variables including age, education, and personality traits. However, none of them improved the model fit to explain the training improvements. While previous research highlighted the effect of these variables on phishing susceptibility or related measures (e.g., [23, 38, 60]), these variables did not seem to impact the proficiency increases and appeared unrelated to personalisation-specific effects. Consistent with [54], the findings suggest that personally identifiable information does not have a substantial influence on training gains.

H1: The training is overall effective. The training was seen as effective and all groups across conditions showed substantial improvements to their phishing proficiency. As confirmed in H1, the personalised groups demonstrated gains in the self-perception of their phishing skills, their security attitude, as well as their behaviour. As the training combined several different methods of phishing detection together (i.e., attachments, links, and biases), participants might have been more able to assess phishing emails more systematically and benefit from insights of methods that fit their needs or complemented their proficiency. For instance, participants can be trained to trigger more systematic thinking based on email cues attempting to exploit cognitive biases (cf. [70]), and then follow this up with learned principles of fake link detection. Such a combination seems promising and should be further evaluated.

H2: Videos are effective, but enhanced with interactivity. Overall, the results demonstrate that the education-only version of the training, consisting solely of the three NoPhish videos [72], can be effective at increasing phishing proficiency. This finding further validates previous work (e.g., [7, 8]) with training that combined all three videos and compared their effect to similar content delivered in other forms. However, participants in the education-only condition showed smaller improvements as compared to the full training for most variables, as predicted by H2. While pattern did not hold for the ability self-estimate, we assume this is due to the variable showing smaller improvements in the training across conditions, pointing to an effect too subtle for our study to detect reliably.

H3: The training levels the playing field. The training was designed to improve participants' phishing proficiency to a similar level, as outlined in the security learning curve [52]. As such, the training provided content that primarily delivered broad foundational improvements in practical skills or knowledge and awareness, rather than honing speciality skills. The results confirm this effect, while still providing improvements for high-proficiency users. Accordingly, our training is effective at targeting users from different backgrounds and bringing them all to an improved, comparable proficiency level.

Phishing and benign detection ability. The phishing detection ability increased substantially between the pre- and post-training classification task, while the benign email detection ability stayed relatively constant. This is to be expected; while the training aimed to increase participants' behaviour by increasing their security behaviour, the goal was not to also increase the ability to identify benign emails as not being phishing. However, the absence of a decrease suggests that the training actually increases users' ability to differentiate phishing emails from benign emails, rather than just raising suspicion in general. This is an important comparison to account for, as some previous phishing training led to unwanted effects where people's suspicion towards benign emails increased as well, leading to overprotective behaviour (e.g., [56]).

Table 9: Summary of hypotheses and related findings.

Hypothesis	Findings
H1: Improvements in Personalised Groups	Consistent training improvements in personalised groups for all variables, except benign email detection accuracy.
H2: Full Training vs. Education-Only	Full training is generally more effective, except for benign email detection accuracy and self-estimates of phishing detection ability.
H3: Increases of Lower Baseline vs. Higher Baseline Proficiency	Participants with a lower baseline proficiency show consistently higher improvements, except for benign email detection accuracy.
H4: Personalised vs. Random Assignment	Mean proficiency was consistently higher in the personalised condition, but the difference from the random assignment condition was not significant.
H5 & H5a: Preferences for Training Assignment and Type	Participants preferred personalisation over random assignment and preferred training that fit their proficiency group compared to others.

Interestingly, a more detailed look at the proficiency groups between the personalised and random assignment condition reveals some different trends. While the benign email detection in the personalised email detection stays relatively similar, there is a tendency for it to decrease in the random assignment condition. Furthermore, small increases in the mean benign detection ability of the low proficiency group actually hints at a slight improvement in the personalised training condition. The different trends between condition diverge even more for the phishing detection accuracy. While all three proficiency groups show almost identical post-training accuracies, it is much more varied in the random assignment condition.

5.2 Benefits of Personalisation

Overall, the results indicate that personalisation appears to be a useful technique in ensuring that training matches proficiency needs, which in turn increases training outcomes. Personalisation of training therefore allows to tailor content more specifically to users' security proficiency, lowering the need for excessive training, and can be effective even with sparse data. Since industry cybersecurity training has limited time and resources available, personalisation can ensure that they are more appropriately assigned to employees based on overall need.

H4: Personalisation seems to be helpful. While participants in the personalised group demonstrated the highest proficiency values and the confidence intervals only narrowly included zero, differences between the personalised and the randomised group were not significant. However, the study might not have had sufficient power to detect more nuanced effects in this comparison, as one third of the participants in the random assignment condition had an identical experience as those in the personalised training condition. Furthermore, participants in the high proficiency group showed lower increases, further limiting the variation that can be explained by the condition differences. Despite these limitations, the results still indicate a trend that the personalisation seems to be effective, with low proficiency participants showing the highest proficiency increases in the personalised training condition. This effect is consistent with findings in learning sciences, which found that children with lower base knowledge benefited more from the personalisation [62].

To account for the indirect measurement of whether training matches participants' needs, the exploratory analysis provides important context. It reveals that training that matches users' proficiency levels enhances total proficiency, phishing detection, self-estimated knowledge and ability, as well as security attitudes. These findings confirm the assumption that the original analysis of personalisation suffered from a lack of power to detect the indirect effect on training effectiveness, but which can be captured if the "matching" factor is directly analysed. Therefore, personalisation appears useful in increasing training effectiveness, beyond already enhancing efficiency, since it assigns users to training that matches their proficiency level.

Previous phishing studies have consistently demonstrated smaller sub groups of users who show either very strong or very poor cybersecurity behaviour (e.g., [35, 60]). Personalisation might be especially effective at addressing these populations, as standard training might not effectively take into account substantial gaps or already existing proficiency.

H5: Participants prefer personalisation. Our results demonstrate a clear preference among participants for personalised phishing training assignments. Many participants favoured personalisation but only as a recommendation while retaining the final choice in training selection. This suggests that, while people value tailored guidance, they also appreciate autonomy in the decision-making process. A substantial portion of participants also preferred fully automated personalisation, which underscores the effectiveness of removing the cognitive load of decision-making and instead trusting the training assignment. This effect was most visible when participants experienced a training that matched their proficiency group, highlighting the importance of experiences with personalisation affecting future willingness to engage with it again.

Cybersecurity campaigns can lead to lower happiness and, especially if more extensive, to substantial time costs. Personalising it can help mitigate these issues by making the training content more tailored to employees' proficiency level and thereby more concise. Even just a simple categorisation into base proficiency levels, such as employed in this study, could be helpful at minimising

such negative effects. Furthermore, personalisation could be more extensive or more granular than the training variants we evaluated. While more research is needed to investigate personalisation strategies, emerging AI-driven technologies could substantially enhance these benefits (see e.g., [28]). Additionally, the use of AI could be used to quickly adapt the personalisation to different job profiles or industry positions.

5.3 Benefits of Different Training Material

Interactive quizzes can help users practice skills and information they learned in prior stages. As proposed by the security learning curve, building up more fundamental awareness to practical skills and finally to employ nudges and reminders to keep a high level of alertness are crucial towards ensuring secure behaviour. Many participants directly hinted at this effect, such as describing how the interactive quizzes were helpful at practising skills and information they learned in earlier stages, while the banners functioned as useful, non-intrusive reminders. However, if the training material does not match users' needs, the resulting frustration can lower engagement with and therefore effectiveness of the training. For instance, a number of participants reported a lack of interest and irritation at the slow pace, or confusion at the more high-level banner information. These proficiency-based differences could explain why the reception to our reminders was more mixed compared to other recent findings, which demonstrated a preference towards non-intrusive reminders [34].

Still, even a basic experience like the education-only condition seemed to provide a good user experience. However, while participants stated that they enjoyed the videos, many participants also complained about a lack of interaction, and that the videos seemed too slow for some; factors that could particularly effect participants with a higher proficiency for whom the videos do not deliver a useful learning experience.

5.4 Recommendations for Phishing Training

Training content. We propose a number of recommendations on how the effectiveness of phishing training could be increased, based on the analysis of our data and related work.

- **Combine multiple phishing detection methods.** Phishing emails use a variety of techniques to manipulate users, from the abuse of cognitive biases to tricking users with fake links and hidden malware in attachments. Training could provide a wider range of content and methods, such as instructional videos to convey self-contained information or interactive quizzes to train and solidify existing knowledge. This would allow users to benefit from training forms that best fit their needs and capabilities.
- **Incorporate interactive elements.** Include interactive quizzes and practical exercises to reinforce knowledge gained in earlier stages of training. This improves skill retention and keeps participants engaged. Participants strongly prefer feedback, which can help not only in phishing detection, but could also hone the skills for benign email detection accuracy.
- **Tailor pace to proficiency.** Adjust the speed and complexity of training materials to match the participant's proficiency level. This helps in making the training more relevant and avoids frustration or boredom, ensuring content meets the user's current skills and needs.

User assignment. Furthermore, we propose a number of recommendations on how users should be assigned to training and what factors are important to factor into that decision.

- **Use a proficiency-based approach.** Proficiency is easy to measure and directly related to training outcomes, thereby easy to justify. The use of proficiency does not require capturing other more sensitive information and can be used to infer the appropriate training variant based on frameworks like the security learning curve [52].
- **Match training to proficiency.** Training can be more efficient overall by assuming different degrees of proficiency and only providing more detailed or fundamental content when necessary. Covering information a person is already familiar with might not only be pointless, but could also cause frustration or disinterest while using time and resources.
- **Offer personalised guidance with autonomy** Provide tailored training recommendations based on individual needs but allow participants the option to choose their own path for a sense of control. If the personalisation is justified well, participants would likely prefer the training that matches their own appropriate proficiency level.

5.5 Limitations and Future Work

This work has three main limitations, (1) related to the sample, (2) the design of the training, and (3) the lack of long-term insights.

First, as this study recruited English-speaking participants from Europe and the United States, the sample is biased towards a western population and the English language. Phishing messages are often not in English [59] and people in different cultures seem to process them differently [21, 65]. Therefore, additional research should explore whether our findings extend to other contexts. Additionally, the training is mostly intended for the use in corporate settings, however, the study questionnaire and setting did not exactly mirror such a context. While we aimed to mimic real-life scenarios with our training as best as possible, participants were still aware of the purpose of the study and the research context. Moreover, as this study employed an online within-between subject study, some conditions had lower participant number than would have been ideal for appropriate power in the statistical analysis.

Second, the analysis showed that the training was most beneficial for low proficiency participants. Future research could look into how personalised training could additionally provide value to higher proficiency participants. As participants also noted in the training feedback, phishing extends beyond the desktop screen and to other contexts like mobile devices. These different contexts affect how we deal with phishing emails [69], which provides further avenues for personalising phishing training to empower users more comprehensively. Furthermore, personalisation could be a promising avenue to account for other inter-individual differences more effectively, such as visual impairment [30].

Third, training effects, especially for lower proficiency participants were promising. However, we are lacking long-term insights into how well the increased proficiency and alertness are retained over time.

Ultimately, the primary goal of the present work was to assess the overall effectiveness of the personalisation effort, and therefore modelled its analysis on the basic experimental setup. Future research could aim to account for greater variability in the data by exploring additional factors, potentially uncovering mediators that influence training outcomes and clarify individual differences in more detail.

6 Conclusion

We evaluated a personalised online phishing training with 342 participants. As part of the personalisation, we evaluated participants' proficiency and assigned them to a training variant that accounted for their prior expertise to achieve a similar post-training proficiency. Participants in lower proficiency groups benefited the most from this personalisation, showing the greatest improvements in proficiency and achieving post-training phishing detection accuracy equivalent to that of high-proficiency participants. We compared this personalisation to random assignment, revealing a tendency towards higher scores in the personalised condition but without confirming a significant overall difference. Nevertheless, the exploratory analysis demonstrated greater effectiveness of proficiency-matched training, which is facilitated by personalisation. Training feedback further revealed a strong preference for personalisation and complementary effects of interactive quizzes following educational videos. We encourage further research, particularly to investigate personalisation techniques that are more granular and that also focus on increasing the expertise of high-proficiency users. However, personalisation appears to be effective in categorising users into a few clearly defined proficiency groups and tailoring training content to those groups. As a result, training with simple personalisation can be more efficient overall, as it better aligns content and delivery with user needs.

7 Data Availability Statement

The data that support the findings of this article are openly available in https://doi.org/10.3929/ethz-b-000721196.

Acknowledgments

This work is graciously supported by armasuisse Science and Technology. We thank Victor Carles for his contribution to a related pilot study [54]. We thank Adrienn Toth for her assistance in drafting and proofreading the manuscript.

References

[1] Hussain Aldawood and Geoffrey Skinner. 2019. Reviewing Cyber Security Social Engineering Training and Awareness Programs—Pitfalls and Ongoing Issues. *Future Internet* 11, 3 (March 2019), 73. https://doi.org/10.3390/fi11030073

[2] S. Alotaibi, Steven Furnell, and Y. He. 2023. Towards a Framework for the Personalization of Cybersecurity Awareness. In *Human Aspects of Information Security and Assurance (IFIP Advances in Information and Communication Technology)*, Steven Furnell and Nathan Clarke (Eds.). Springer Nature Switzerland, Cham, Switzerland, 143–153. https://doi.org/10.1007/978-3-031-38530-8_12

[3] APA. 2017. Ethical principles of psychologists and code of conduct. https://www.apa.org/ethics/code

[4] APWG. 2023. Phishing Activity Trends Report, 1st Quarter 2023. https://docs.apwg.org/reports/apwg_trends_report_q1_2023.pdf

[5] Steffen Bartsch and Melanie Volkamer. 2012. Towards the Systematic Development of Contextualized Security Interventions. In *Designing Interactive Secure Systems: Workshop at British HCI 2012, University of Birmingham, 11th September 2012.* University of Birmingham, Birmingham, UK, 1–4.

[6] Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67, 1 (2015), 1–48. https://doi.org/10.18637/jss.v067.i01

[7] Benjamin M. Berens, Mattia Mossano, and Melanie Volkamer. 2024. Taking 5 minutes protects you for 5 months: Evaluating an anti-phishing awareness video. *Computers & Security* 137 (Feb. 2024), 103620. https://doi.org/10.1016/j.cose.2023.103620

[8] Benjamin Maximilian Berens, Florian Schaub, Mattia Mossano, and Melanie Volkamer. 2024. Better Together: The Interplay Between a Phishing Awareness Video and a Link-centric Phishing Support Tool. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (New York, NY, USA). Association for Computing Machinery, New York, NY, USA, 1–60. https://doi.org/10.1145/3613904.3642843

[9] Matthew L Bernacki, Meghan J Greene, and Nikki G Lobczowski. 2021. A systematic review of research on personalized learning: Personalized by whom, to what, how, and for what purpose (s)? *Educational Psychology Review* 33, 4 (2021), 1675–1715.

[10] Marcus Beyer, Sarah Ahmed, Katja Doerlemann, Simon Arnell, Simon Parkin, Angela Sasse, and Neil Passingham. 2016. Awareness is only the first step: A framework for progressive engagement of staff in cyber security.

[11] Aditi Bhutoria. 2022. Personalized education and Artificial Intelligence in the United States, China, and India: systematicreview using a Human-In-The-Loop model. *Computers and Education: Artificial Intelligence* 3 (Jan. 2022), 100068. https://doi.org/10.1016/j.caeai.2022.100068

[12] Marcus Butavicius, Kathryn Parsons, Malcolm Pattinson, and Agata McCormac. 2016. Breaching the Human Firewall: Social engineering in Phishing and Spear-Phishing Emails. https://doi.org/10.48550/arXiv.1606.00887

[13] Deanna D. Caputo, Shari Lawrence Pfleeger, Jesse D. Freeman, and M. Eric Johnson. 2014. Going Spear Phishing: Exploring Embedded Training and Awareness. *IEEE Security & Privacy* 12, 1 (Jan. 2014), 28–38. https://doi.org/10.1109/MSP.2013.106

[14] Sherry Y. Chen and Jen-Han Wang. 2021. Individual differences and personalized learning: a review and appraisal. *Universal Access in the Information Society* 20, 4 (Nov. 2021), 833–849. https://doi.org/10.1007/s10209-020-00753-4

[15] Nabin Chowdhury and Vasileios Gkioulos. 2021. Cyber security training for critical infrastructure protection: A literature review. *Computer Science Review* 40 (May 2021), 100361. https://doi.org/10.1016/j.cosrev.2021.100361

[16] Clayton R. Cook, Stephen P. Kilgus, and Matthew K. Burns. 2018. Advancing the science and practice of precision education to enhance student outcomes. *Journal of School Psychology* 66 (Feb. 2018), 4–10. https://doi.org/10.1016/j.jsp.2017.11.004

[17] Yuli Deng, Duo Lu, Chun-Jen Chung, Dijiang Huang, and Zhen Zeng. 2018. Personalized Learning in a Virtual Hands-on Lab Platform for Computer Science Education. In *2018 IEEE Frontiers in Education Conference (FIE).* IEEE, New York, NY, USA, 1–8. https://doi.org/10.1109/FIE.2018.8659291

[18] Serge Egelman, Lorrie Faith Cranor, and Jason Hong. 2008. You've been warned: an empirical study of the effectiveness of web browser phishing warnings. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* ACM, Florence Italy, 1065–1074. https://doi.org/10.1145/1357054.1357219

[19] Cori Faklaris, Laura Dabbish, and Jason I. Hong. 2022. Do They Accept or Resist Cybersecurity Measures? Development and Validation of the 13-Item Security Attitude Inventory (SA-13). https://doi.org/10.48550/arXiv.2204.03114

[20] Rida Indah Fariani, Kasiyah Junus, and Harry Budi Santoso. 2023. A Systematic Literature Review on Personalised Learning in the Higher Education Context. *Technology, Knowledge and Learning* 28, 2 (June 2023), 449–476. https://doi.org/10.1007/s10758-022-09628-4

[21] Waldo Rocha Flores, Hannes Holm, Marcus Nohlberg, and Mathias Ekstedt. 2015. Investigating personal determinants of phishing and the effect of national culture. *Information & Computer Security* 23, 2 (June 2015), 178–199. https://doi.org/10.1108/ICS-05-2014-0029

[22] Anjuli Franz, Verena Zimmermann, Gregor Albrecht, Katrin Hartwig, Christian Reuter, Alexander Benlian, and Joachim Vogt. 2021. SoK: Still Plenty of Phish in the Sea - A Taxonomy of User-Oriented Phishing Interventions and Avenues for Future Research. In *17th Symposium on Usable Privacy and Security (SOUPS 2021), Online, August 8-10, 2021.* USENIX Association, Berkeley, CA, USA, 358. https://www.usenix.org/conference/soups2021/presentation/franz

[23] Edwin Donald Frauenstein and Stephen Flowerday. 2020. Susceptibility to phishing on social network sites: A personality information processing model. *Computers & Security* 94 (2020), 101862. https://doi.org/10.1016/j.cose.2020.101862

[24] Sumair Ijaz Hashmi, Niklas George, Eimaan Saqib, Fatima Ali, Nawaal Siddique, Shafay Kashif, Shahzaib Ali, Nida Ul Habib Bajwa, and Mobin Javed. 2023. Training Users to Recognize Persuasion Techniques in Vishing Calls. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems (CHI EA '23).* Association for Computing Machinery, New York, NY, USA, 1–8. https://doi.org/10.1145/3544549.3585823

[25] Wu He and Zuopeng (Justin) Zhang. 2019. Enterprise cybersecurity training and awareness programs: Recommendations for success. *Journal of Organizational Computing and Electronic Commerce* 29, 4 (Oct. 2019), 249–257. https://doi.org/10.1080/10919392.2019.1611528

[26] Jonas Hielscher, Annette Kluge, Uta Menges, and M. Angela Sasse. 2022. "Taking out the trash": Why security behavior change requires intentional forgetting. In *Proceedings of the 2021 New Security Paradigms Workshop*. Association for Computing Machinery, New York, NY, USA, 108–122. https://doi.org/10.1145/3498891.3498902

[27] Shiu-Li Huang and Jung-Hung Shiu. 2012. A user-centric adaptive learning system for e-learning 2.0. *Journal of Educational Technology & Society* 15, 3 (2012), 214–225.

[28] Farnaz Jahanbakhsh, Yannis Katsis, Dakuo Wang, Lucian Popa, and Michael Muller. 2023. Exploring the Use of Personalized AI for Identifying Misinformation on Social Media. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 105, 27 pages. https://doi.org/10.1145/3544548.3581219

[29] Daniel Jampen, Gürkan Gür, Thomas Sutter, and Bernhard Tellenbach. 2020. Don't click: towards an effective anti-phishing training. A comparative literature review. *Human-centric Computing and Information Sciences* 10, 1 (Aug. 2020), 33. https://doi.org/10.1186/s13673-020-00237-7

[30] Emaan Bilal Khan, Emaan Atique, and Mobin Javed. 2024. Investigating Phishing Threats in the Email Browsing Experience of Visually Impaired Individuals. In *Extended Abstracts of the 2024 CHI Conference on Human Factors in Computing Systems (CHI EA '24)*. Association for Computing Machinery, New York, NY, USA, Article 206, 11 pages. https://doi.org/10.1145/3613905.3651076

[31] Aleksandra Klašnja-Milićević, Boban Vesin, Mirjana Ivanović, and Zoran Budimac. 2011. E-Learning personalization based on hybrid recommendation strategy and learning style identification. *Computers & Education* 56, 3 (April 2011), 885–899. https://doi.org/10.1016/j.compedu.2010.11.001

[32] Ponnurangam Kumaraguru, Justin Cranshaw, Alessandro Acquisti, Lorrie Cranor, Jason Hong, Mary Ann Blair, and Theodore Pham. 2009. School of phish: a real-world evaluation of anti-phishing training. In *Proceedings of the 5th Symposium on Usable Privacy and Security (SOUPS '09)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/1572532.1572536

[33] Ponnurangam Kumaraguru, Steve Sheng, Alessandro Acquisti, Lorrie Faith Cranor, and Jason Hong. 2010. Teaching Johnny not to fall for phish. *ACM Transactions on Internet Technology* 10, 2 (2010), 1–31. https://doi.org/10.1145/1754393.1754396

[34] Daniele Lain, Tarek Jost, Sinisa Matetic, Kari Kostiainen, and Srdjan Capkun. 2024. Content, Nudges and Incentives: A Study on the Effectiveness and Perception of Embedded Phishing Training. https://doi.org/10.1145/3658644.3690348

[35] Daniele Lain, Kari Kostiainen, and Srdjan Capkun. 2022. Phishing in Organizations: Findings from a Large-Scale and Long-Term Study. In *2022 IEEE Symposium on Security and Privacy (SP)*. IEEE, San Francisco, CA, USA, 842–859. https://doi.org/10.1109/sp46214.2022.9833766 Accepted: 2023-01-04T13:37:50Z.

[36] Russell V. Lenth. 2024. *emmeans: Estimated Marginal Means, aka Least-Squares Means*. R Package. https://CRAN.R-project.org/package=emmeans

[37] Daniel Leyzberg, Aditi Ramachandran, and Brian Scassellati. 2018. The Effect of Personalization in Longer-Term Robot Tutoring. *J. Hum.-Robot Interact.* 7, 3 (Dec. 2018), 19:1–19:19. https://doi.org/10.1145/3283453

[38] Pablo López-Aguilar and Agusti Solanas. 2021. Human Susceptibility to Phishing Attacks Based on Personality Traits: The Role of Neuroticism. In *2021 IEEE 45th Annual Computers, Software, and Applications Conference (COMPSAC)*. IEEE, Madrid, Spain, 1363–1368. https://doi.org/10.1109/COMPSAC51774.2021.00192

[39] Louis Major and Gill A. Francis. 2020. *Technology-Supported Personalised Learning: A Rapid Evidence Review*. Technical Report 1. EdTech Hub. https://doi.org/10.5281/zenodo.4556925

[40] Louis Major, Gill A. Francis, and Maria Tsapali. 2021. The effectiveness of technology-supported personalised learning in low- and middle-income countries: A meta-analysis. *British Journal of Educational Technology* 52, 5 (2021), 1935–1964. https://doi.org/10.1111/bjet.13116

[41] Claudio Marforio, Ramya Jayaram Masti, Claudio Soriente, Kari Kostiainen, and Srdjan Čapkun. 2016. Evaluation of Personalized Security Indicators as an Anti-Phishing Mechanism for Smartphone Applications. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 540–551. https://doi.org/10.1145/2858036.2858085

[42] Alexis R. Neigel, Victoria L. Claypoole, Grace E. Waldfogle, Subrata Acharya, and Gabriella M. Hancock. 2020. Holistic cyber hygiene education: Accounting for the human factors. *Computers & Security* 92 (May 2020), 101731. https://doi.org/10.1016/j.cose.2020.101731

[43] Kathryn Parsons, Dragana Calic, Malcolm Pattinson, Marcus Butavicius, Agata McCormac, and Tara Zwaans. 2017. The Human Aspects of Information Security Questionnaire (HAIS-Q): Two further validation studies. *Computers & Security* 66 (May 2017), 40–51. https://doi.org/10.1016/j.cose.2017.01.004

[44] Maria Perez-Ortiz, Claire Dormann, Yvonne Rogers, Sahan Bulathwela, Stefan Kreitmayer, Emine Yilmaz, Richard Noss, and John Shawe-Taylor. 2021. X5Learn: A Personalised Learning Companion at the Intersection of AI and HCI. In *Companion Proceedings of the 26th International Conference on Intelligent User Interfaces* (College Station, TX, USA) *(IUI '21 Companion)*. Association for Computing Machinery, New York, NY, USA, 70–74. https://doi.org/10.1145/3397482.3450721

[45] Muh Putra Pratama, Rigel Sampelolo, and Hans Lura. 2023. Revolutionizing education: harnessing the power of artificial intelligence for personalized learning. *Klasikal: Journal of education, language teaching and science* 5, 2 (2023), 350–357.

[46] R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

[47] Beatrice Rammstedt, Christoph J. Kemper, Mira Céline Klein, Constanze Beierlein, and Anastassiya Kovaleva. 2013. A Short Scale for Assessing the Big Five Dimensions of Personality: 10 Item Big Five Inventory (BFI-10). *methods, data, analyses* 7, 2 (2013), 17. https://doi.org/10.12758/mda.2013.013 Number: 2.

[48] Benjamin Reinheimer, Lukas Aldag, Peter Mayer, Mattia Mossano, Reyhan Duezguen, Bettina Lofthouse, Tatiana von Landesberger, and Melanie Volkamer. 2020. An investigation of phishing awareness and education over time: When and how to best remind users. In *Sixteenth Symposium on Usable Privacy and Security (SOUPS 2020)*. Usenix, Berkeley, CA, USA, 259–284. https://www.usenix.org/conference/soups2020/presentation/reinheimer

[49] Karen Renaud, Melanie Volkamer, and Arne Renkema-Padmos. 2014. Why doesn't Jane protect her privacy?. In *Privacy Enhancing Technologies*, Emiliano De Cristofaro and Steven J. Murdoch (Eds.). Springer International Publishing, Cham, Switzerland, 244–262. https://doi.org/10.1007/978-3-319-08506-7_13

[50] Rene Roepke, Vincent Drury, Ulrike Meyer, and Ulrik Schroeder. 2022. Better the Phish You Know: Evaluating Personalization in Anti-Phishing Learning Games. In *Proceedings of the 14th International Conference on Computer Supported Education*. SCITEPRESS - Science and Technology Publications, Setúbal, Portugal, 458–466. https://doi.org/10.5220/0011042100003182

[51] Orvila Sarker, Asangi Jayatilaka, Sherif Haggag, Chelsea Liu, and M. Ali Babar. 2024. A Multi-vocal Literature Review on challenges and critical success factors of phishing education, training and awareness. *Journal of Systems and Software* 208 (2024), 111899. https://doi.org/10.1016/j.jss.2023.111899

[52] M. Angela Sasse, Jonas Hielscher, Jennifer Friedauer, and Annalina Buckmann. 2023. Rebooting IT security awareness–how organisations can encourage and sustain secure behaviours. In *Computer Security. ESORICS 2022 International Workshops*. Springer International Publishing, Cham, Switzerland, 248–265. https://doi.org/10.1007/978-3-031-25460-4_14

[53] M. A. Sasse, J. Hielscher, J. Friedauer, and A. Buckmann. 2023. Rebooting IT Security Awareness – How Organisations Can Encourage and Sustain Secure Behaviours. https://doi.org/10.1007/978-3-031-25460-4_14 Conference Name: European Symposium on Research in Computer Security ISSN: 0302-9743 Meeting Name: European Symposium on Research in Computer Security Pages: 248-265 Publisher: Springer, Cham Volume: 13785.

[54] Lorin Schöni, Victor Carles, Martin Strohmeier, Peter Mayer, and Verena Zimmermann. 2024. You Know What? - Evaluation of a Personalised Phishing Training Based on Users' Phishing Knowledge and Detection Skills. In *Proceedings of the 2024 European Symposium on Usable Security (EuroUSEC '24)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3688459.3688460

[55] Pavel Seda, Jan Vykopal, Valdemar Švábenský, and Pavel Čeleda. 2021. Reinforcing Cybersecurity Hands-on Training With Adaptive Learning. In *2021 IEEE Frontiers in Education Conference (FIE)*. IEEE, New York, NY, USA, 1–9. https://doi.org/10.1109/FIE49875.2021.9637252

[56] Steve Sheng, Mandy Holbrook, Ponnurangam Kumaraguru, Lorrie Faith Cranor, and Julie Downs. 2010. Who falls for phish? a demographic analysis of phishing susceptibility and effectiveness of interventions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*. Association for Computing Machinery, New York, NY, USA, 373–382. https://doi.org/10.1145/1753326.1753383

[57] Ansar Siddique, Qaiser S Durrani, and Husnain A Naqvi. 2019. Developing adaptive e-learning environment using cognitive and noncognitive parameters. *Journal of Educational Computing Research* 57, 4 (2019), 811–845.

[58] Mario Silic and Andrea Back. 2016. The dark side of social networking sites: Understanding phishing risks. *Computers in Human Behavior* 60 (July 2016), 35–43. https://doi.org/10.1016/j.chb.2016.02.050

[59] Camelia Simoiu, Ali Zand, Kurt Thomas, and Elie Bursztein. 2020. Who is targeted by email-based phishing and malware? Measuring factors that differentiate risk. In *Proceedings of the ACM Internet Measurement Conference (IMC '20)*. Association for Computing Machinery, New York, NY, USA, 567–576. https://doi.org/10.1145/3419394.3423617

[60] Teodor Sommestad and Henrik Karlzén. 2019. A meta-analysis of field experiments on phishing susceptibility. In *2019 APWG Symposium on Electronic Crime Research (eCrime)*. IEEE, Pittsburgh, PA, USA, 1–14. https://doi.org/10.1109/eCrime47957.2019.9037502 ISSN: 2159-1245.

[61] Daniel Stein, Benjamin Scribner, Noel Kyle, William Newhouse, Clarence Williams, and Baris Yakin. 2017. *National Initiative for Cybersecurity Education (NICE) Framework Work Role Capability Indicators: Indicators for Performing Work Roles.* Technical Report NIST Internal or Interagency Report (NISTIR) 8193 (Draft). National Institute of Standards and Technology. https://csrc.nist.gov/pubs/ir/8193/ipd

[62] Klaus D. Stiller and Rosemarie Jedlicka. 2010. A kind of expertise reversal effect: Personalisation effect can depend on domain-specific prior knowledge. *Australasian Journal of Educational Technology* 26, 1 (March 2010), 133–149. https://doi.org/10.14742/ajet.1107 Number: 1.

[63] Soni Sweta and Kanhaiya Lal. 2017. Personalized Adaptive Learner Model in E-Learning System Using FCM and Fuzzy Inference System. *International Journal of Fuzzy Systems* 19, 4 (Aug. 2017), 1249–1260. https://doi.org/10.1007/s40815-017-0309-y

[64] Olga Tapalova and Nadezhda Zhiyenbayeva. 2022. Artificial intelligence in education: AIEd for personalised learning pathways. *Electronic Journal of e-Learning* 20, 5 (2022), 639–653.

[65] Rucha Tembe, Olga Zielinska, Yuqi Liu, Kyung Wha Hong, Emerson Murphy-Hill, Chris Mayhorn, and Xi Ge. 2014. Phishing in international waters: exploring cross-national differences in phishing conceptualizations between Chinese, Indian and American samples. In *Proceedings of the 2014 Symposium and Bootcamp on the Science of Security (HotSoS '14).* Association for Computing Machinery, New York, NY, USA, 1–7. https://doi.org/10.1145/2600176.2600178

[66] Rani Van Schoors, Jan Elen, Annelies Raes, and Fien Depaepe. 2021. An overview of 25 years of research on digital personalised learning in primary and secondary education: A systematic review of conceptual and methodological trends. *British Journal of Educational Technology* 52, 5 (2021), 1798–1822.

[67] Ismini Vasileiou and Steven Furnell. 2023. Enhancing Security Education - Recognising Threshold Concepts and Other Influencing Factors. In *Proceedings of the 4th International Conference on Information Systems Security and Privacy - ICISSP*. SciTePress, Funchal, Madeira, 398–403. https://doi.org/10.5220/0006646203980403

[68] Laton Vermette, Joanna McGrenere, Colin Birge, Adam Kelly, and Parmit K. Chilana. 2019. Freedom to Personalize My Digital Classroom: Understanding Teachers' Practices and Motivations. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19).* Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3290605.3300548

[69] Arun Vishwanath. 2016. Mobile device affordance: Explicating how smartphones influence the outcome of phishing attacks. *Computers in Human Behavior* 63 (2016), 198–207. https://doi.org/10.1016/j.chb.2016.05.035

[70] Arun Vishwanath, Brynne Harrison, and Yu Jie Ng. 2018. Suspicion, Cognition, and Automaticity Model of Phishing Susceptibility. *Communication Research* 45, 8 (2018), 1146–1166. https://doi.org/10.1177/0093650215627483

[71] Arun Vishwanath, Tejaswini Herath, Rui Chen, Jingguo Wang, and H. Raghav Rao. 2011. Why do people get phished? Testing individual differences in phishing vulnerability within an integrated, information processing model. *Decision Support Systems* 51, 3 (2011), 576–586. https://doi.org/10.1016/j.dss.2011.03.002

[72] Melanie Volkamer, Karen Renaud, Benjamin Reinheimer, Philipp Rack, Marco Ghiglieri, Peter Mayer, Alexandra Kunz, and Nina Gerber. 2018. Developing and Evaluating a Five Minute Phishing Awareness Video. In *Trust, Privacy and Security in Digital Business* (Cham), Steven Furnell, Haralambos Mouratidis, and Günther Pernul (Eds.). Springer International Publishing, Regensburg, Germany, 119–134. https://doi.org/10.1007/978-3-319-98385-1_9

[73] Rick Wash. 2020. How Experts Detect Phishing Scam Emails. *Proceedings of the ACM on Human-Computer Interaction* 4 (2020), 160:1–160:28. Issue CSCW2. https://doi.org/10.1145/3415231

[74] Zikai Alex Wen, Zhiqiu Lin, Rowena Chen, and Erik Andersen. 2019. What.Hack: Engaging Anti-Phishing Training Through a Role-playing Phishing Simulation Game. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19).* Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300338

[75] Eva Wolfangel. 2023. The Human Element in Cybercrime and Cybersecurity. https://www.youtube.com/watch?v=LKUMRTLV49g

[76] Shouhuai Xu. 2019. Cybersecurity Dynamics: A Foundation for the Science of Cybersecurity. In *Proactive and Dynamic Network Defense*, Cliff Wang and Zhuo Lu (Eds.). Springer International Publishing, Cham, 1–31. https://doi.org/10.1007/978-3-030-10597-6_1

[77] Beste F. Yuksel, Kurt B. Oleson, Lane Harrison, Evan M. Peck, Daniel Afergan, Remco Chang, and Robert JK Jacob. 2016. Learn Piano with BACh: An Adaptive Learning Interface that Adjusts Task Difficulty Based on Brain State. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16).* Association for Computing Machinery, New York, NY, USA, 5372–5384. https://doi.org/10.1145/2858036.2858388

[78] Fatemeh Mariam Zahedi, Yan Chen, and Huimin Zhao. 2023. Ontology-Based Intelligent Interface Personalization for Protection Against Phishing Attacks. *Information Systems Research* 35, 3 (Oct. 2023), 1463–1478. https://doi.org/10.1287/isre.2021.0065

[79] Markus Zanker, Laurens Rook, and Dietmar Jannach. 2019. Measuring the impact of online personalisation: Past, present and future. *International Journal of Human-Computer Studies* 131 (Nov. 2019), 160–168. https://doi.org/10.1016/j.ijhcs.2019.06.006

[80] Verena Zimmermann and Karen Renaud. 2019. Moving from a 'human-as-problem" to a 'human-as-solution" cybersecurity mindset. *International Journal of Human-Computer Studies* 131 (Nov. 2019), 169–187. https://doi.org/10.1016/j.ijhcs.2019.05.005

A Background Task

We implemented a background task that loosely mirrors the mailbox of an employee working at a company, following a scenario where the employee just came back to work from a holiday and now has to keep up with various emails. These emails include work-related topics, such as discussing tasks and news, sending meeting minutes, or scheduling appointments. Furthermore, some of the emails are not strictly work-related, such as asking if anyone found a lost scarf or people to an event. Finally, some emails represent spam or phishing. We provide the full source code of the background task in the overall training code repository on GitHub.[3] Example screenshots of the background task emails can be seen in Fig. 11.

The background task represented the base activity during the training period. Participants in the personalised and randomised control conditions were redirected to this background task immediately after completing the pre-training questionnaires. After the training period, participants were again immediately redirected to the post-training questionnaires. Participants in the education-only control condition did not experience the background task.

[3] GitHub Repository: https://github.com/lorinschoeni/personalised-phishing

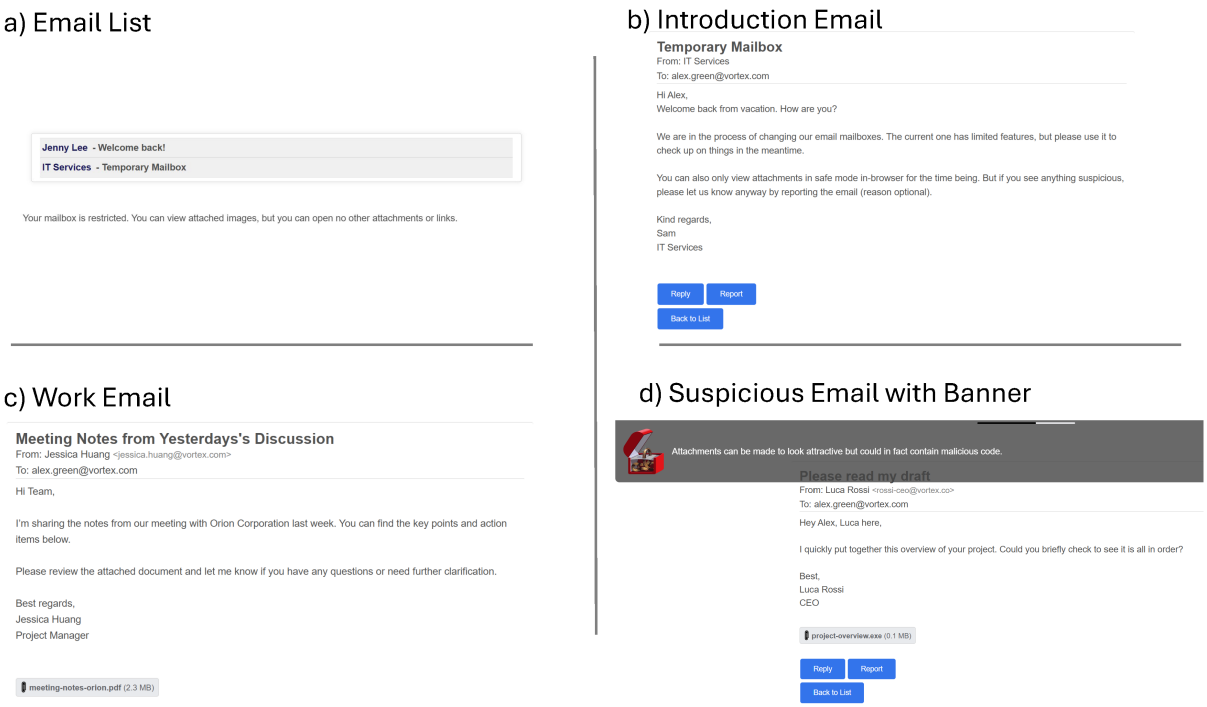


Figure 11: Overview of the background task that included a) the landing screen and email list with the first emails participants saw, b) the first introductory email that participants would have seen, c) an example of a work email seen later in the task, and d) an example of a suspicious email with a reminder banner superimposed at the top. Participants could freely move between emails as they came in, and reply to them or report and delete them.

B Statistical Results: Model Fit Comparisons & Regression Models

Control variables. We evaluated whether the pre-registered control variables could explain additional variability in the data beyond general factors of the experimental paradigm. To assess whether any of the collected control variables had an influence on the intervention effect, we compared whether separate models containing each of the control variables could provide a better fit than our base model. We used the total proficiency base model (see column 1 in Table 12) as it represented the general increase affected by other measures, and any substantial effect would therefore be reflected in its estimates. We provide a summary of these comparisons in Table 10. As the models were nested, a simple ANOVA comparison was suitable.

Training match. We additionally conducted an exploratory analysis of whether the training's proficiency level matched the participants' proficiency level, which provided a more accurate comparison of the personalisation benefit. We provide a summary of this comparison in Table 11. As the models were nested, a simple ANOVA comparison was suitable.

Regression tables. We conducted regression analysis for each of our variables of interest, to evaluate how they are influenced by the training and experimental setup. We provide summary tables of the resulting regression models (Table 12 and Table 13) below.

Table 10: Results of the ANOVA comparison between the base proficiency model and models with control variables. No models provided a significantly enhanced fit.

Model	Res.Df	RSS	Df	Sum of Sq	F-value	Pr(>F)
Base Proficiency Model	666	3823.2	-	-	-	-
Addition of Age	661	3798.7	5	24.587	0.856	0.511
Addition of Education	661	3810.1	5	13.14	0.456	0.809
Addition of Agreeableness	665	3809.5	1	13.772	2.404	0.122
Addition of Extraversion	665	3817.6	1	5.666	0.987	0.321
Addition of Conscientiousness	665	3820.1	1	3.111	0.542	0.462
Addition of Neuroticism	665	3814.2	1	9.071	1.582	0.209
Addition of Openness	665	3811.1	1	12.156	2.121	0.146

Table 11: Results of the ANOVA comparison between the base proficiency model and a model with a binary variable of whether training matches users' proficiency level. The model with the additional training match variable provided a significantly enhanced fit.

Model	Res.Df	RSS	Df	Sum of Sq	F-value	Pr(>F)
Base Proficiency Model	666	3823.2	-	-	-	-
Addition of Training Match	664	3768.0	2	55.22	4.865	0.008

Table 12: Summary of Regression Models

	Dependent variable:			
	Total Proficiency	Phish Detection Accuracy	Benign Detection Accuracy	SA-13 Total
	(1)	(2)	(3)	(4)
timepre	−4.820***	−0.257***	−0.017	−0.427***
	(0.573)	(0.047)	(0.046)	(0.127)
condition1	−1.631**	−0.056	−0.071	0.188
	(0.635)	(0.052)	(0.051)	(0.141)
condition2	−0.562	−0.101*	−0.043	0.089
	(0.643)	(0.052)	(0.052)	(0.143)
group1	0.934*	0.009	0.072*	0.185*
	(0.493)	(0.040)	(0.040)	(0.110)
group2	3.120***	0.002	0.144***	0.407***
	(0.528)	(0.043)	(0.042)	(0.117)
timepre:condition1	1.316	0.035	0.082	0.010
	(0.898)	(0.073)	(0.072)	(0.200)
timepre:condition2	0.507	0.033	0.075	−0.008
	(0.910)	(0.074)	(0.073)	(0.202)
timepre:group1	2.580***	0.072	0.024	0.202
	(0.697)	(0.057)	(0.056)	(0.155)
timepre:group2	3.970***	0.187***	0.027	0.363**
	(0.747)	(0.061)	(0.060)	(0.166)
condition1:group1	1.256	0.008	−0.023	−0.084
	(0.779)	(0.063)	(0.063)	(0.173)
condition2:group1	0.507	0.026	−0.026	−0.039
	(0.802)	(0.065)	(0.065)	(0.178)
condition1:group2	1.829**	0.079	0.013	−0.071
	(0.865)	(0.070)	(0.070)	(0.192)
condition2:group2	0.711	0.085	−0.092	0.176
	(0.857)	(0.070)	(0.069)	(0.191)
timepre:condition1:group1	−0.830	−0.009	0.054	−0.111
	(1.102)	(0.090)	(0.089)	(0.245)
timepre:condition2:group1	0.133	0.091	0.020	−0.014
	(1.135)	(0.092)	(0.091)	(0.252)
timepre:condition1:group2	−1.326	−0.032	−0.065	−0.096
	(1.223)	(0.100)	(0.098)	(0.272)
timepre:condition2:group2	−0.332	−0.010	−0.027	−0.093
	(1.211)	(0.099)	(0.097)	(0.269)
Constant	21.906***	0.848***	0.781***	3.556***
	(0.405)	(0.033)	(0.033)	(0.090)
Observations	684	684	684	684
R^2	0.517	0.218	0.102	0.207
Adjusted R^2	0.504	0.198	0.079	0.187

*p<0.1; **p<0.05; ***p<0.01

Table 13: Summary of Self-Estimate Regression Models

	Dependent variable:		
	Self-Estimated Knowledge	Self-Estimated Ability	Self-Estimated Alertness
	(1)	(2)	(3)
timepre	−1.371***	−0.886***	−1.286***
	(0.175)	(0.169)	(0.178)
condition1	−0.357*	−0.373**	−0.279
	(0.194)	(0.187)	(0.198)
condition2	−0.161	−0.001	−0.376*
	(0.197)	(0.190)	(0.200)
group1	−0.145	−0.038	−0.152
	(0.151)	(0.145)	(0.153)
group2	0.203	0.186	0.111
	(0.162)	(0.156)	(0.164)
timepre:condition1	0.371	0.094	0.119
	(0.275)	(0.265)	(0.280)
timepre:condition2	0.545*	0.147	0.503*
	(0.278)	(0.268)	(0.283)
timepre:group1	0.700***	0.379*	0.587***
	(0.213)	(0.205)	(0.217)
timepre:group2	1.231***	0.846***	1.086***
	(0.228)	(0.220)	(0.232)
condition1:group1	0.514**	0.300	0.424*
	(0.238)	(0.230)	(0.242)
condition2:group1	0.265	0.098	0.631**
	(0.245)	(0.237)	(0.250)
condition1:group2	0.297	0.393	0.499*
	(0.265)	(0.255)	(0.269)
condition2:group2	0.351	0.187	0.701***
	(0.262)	(0.253)	(0.267)
timepre:condition1:group1	−0.439	0.043	−0.029
	(0.337)	(0.325)	(0.343)
timepre:condition2:group1	−0.743**	−0.298	−0.462
	(0.347)	(0.335)	(0.353)
timepre:condition1:group2	−0.311	−0.134	−0.239
	(0.374)	(0.361)	(0.381)
timepre:condition2:group2	−0.584	−0.071	−0.482
	(0.371)	(0.357)	(0.377)
Constant	3.857***	3.914***	4.029***
	(0.124)	(0.119)	(0.126)
Observations	684	684	684
R^2	0.335	0.251	0.319
Adjusted R^2	0.318	0.231	0.302

*p<0.1; **p<0.05; ***p<0.01