MIS Quarterly

RESEARCH NOTE

LEARNING BY PHISHING VIA POST-SIMULATION FEEDBACK: FROM EMBEDDED TO NON-EMBEDDED TRAINING[1]

Dezhi Yin and Matthew Mullarkey
Muma College of Business, University of South Florida
Tampa, FL, U.S.A. {dezhiyin@usf.edu} {mmullarkey@usf.edu}

Gert-Jan de Vreede
School of Business, Stevens Institute of Technology
Hoboken, NJ, U.S.A. {gj@stevens.edu}

Moez Limayem
University of South Florida
Tampa, FL, U.S.A. {mlimayem@usf.edu}

Given the frequent occurrence of phishing attacks and their devastating consequences, organizations are increasingly deploying phishing simulation emails to quantify employee susceptibility (e.g., clicking within-email links) and investing in training programs to reduce such susceptibility. Interestingly, phishing simulations can also be turned into a training opportunity in themselves. A best practice in industry is "embedded training"—providing immediate feedback on landing pages to employees who fail the simulations. This intervention is intuitively appealing given its "just-in-time" nature. Although laboratory studies from the literature have offered broad support for its effectiveness in reducing employee susceptibility, studies conducted in field settings have observed weaker evidence or even a reversed effect that increased susceptibility. In this research, we recognize an inherent shortcoming of the real-world implementation of embedded training: limited reach. To address this practical challenge, we propose an alternative, novel intervention—"non-embedded training"—that decouples feedback from the failure action and sends delayed feedback to all the employees. Following an "empirics-first" approach, we conducted three randomized field experiments using a leading phishing simulation platform to explore the respective and combined effects of embedded and non-embedded training in reducing user vulnerability over time. This research contributes to the practice and literature on phishing and cybersecurity by challenging the assumed effectiveness of embedded training in practice and revealing how non-embedded training could be a more promising intervention.

Keywords: Phishing attacks, social engineering, cybersecurity, phishing simulations, embedded training, non-embedded training, timing, exposure

Introduction

A widespread and dangerous type of cybersecurity risk for organizations is the phishing attack, whereby cybercriminals send emails or other types of messages to employees and ask them to click a link or open an attachment. The emails often mimic the look and feel of legitimate ones and use persuasion techniques to deceive the recipients and increase click-through rates (Wright et al., 2014). Phishing attacks are likely to grow more sophisticated with the rise of generative AI tools (e.g., FraudGPT) that facilitate the generation of authentic-looking phishing emails. Once employees fall for such an

[1] John D'Arcy was the accepting senior editor for this paper. Markus Weinmann served as the associate editor. The transparency materials for this article can be found at https://osf.io/ut672/

attack, they may expose sensitive information or install malware, which can then be used for financial fraud, identity theft, and other crimes.

Organizations have different means to address the risk of phishing attacks. They can detect phishing attempts by implementing technological solutions based on web address legitimacy, webpage content similarity, etc. (Abbasi et al., 2015; Vijayalakshmi et al., 2020). However, technological solutions alone are rarely sufficient at preventing end-user exposure, and a single susceptible employee clicking on a well-engineered phishing email can compromise the entire security system. Humans are often the weakest point of the system and the last defense against phishing attacks. Thus, companies are increasingly deploying phishing simulations—fabricated emails or messages that mimic real-world phishing attempts but pose no actual danger—to identify the most susceptible employees and then ask such employees to complete relevant training programs (Jensen et al., 2017, 2022).

Apart from traditional training programs, phishing simulations themselves can be turned into a training opportunity—the focus of our research. A "best practice" in industry is called "embedded training" (Cyberready.com, 2025). Embedded training takes advantage of a phishing simulation and delivers *immediate* feedback on the landing page once a user clicks on the link and *fails* the simulation. Embedded training is brief and efficient by design because the landing page can provide example-based feedback as simple as a list of "red flags" that users may have missed earlier (Pan et al., 2024). As such, embedded training often requires less user commitment than traditional training programs. For example, the default and most popular landing page for a phishing campaign on KnowBe4, a security awareness training and simulated phishing platform, is "Oops!" followed by some tips (KnowBe4, 2025a).

A premise of embedded training is that the most "teachable moment" is the exact moment of failure (Franz et al., 2021)[2]. A theoretical basis behind this premise originates from Skinner's reinforcement theory. Based on this theory, learning occurs through feedback, which should be given *immediately* after certain actions to eliminate incorrect actions and reinforce correct ones (Skinner, 1954). Laboratory studies using scenario-based stimuli have generally found that embedded training enhances users' accuracy in distinguishing phishing from legitimate emails (e.g., Kumaraguru et al., 2007, 2010; Pan et al., 2024). However, field studies carrying out real-world implementations of embedded training have revealed limited evidence for the effectiveness of embedded training in reducing actual user vulnerability captured in subsequent phishing tests (e.g., Ho et al., 2025; Lain et al., 2022). For example, Lain et al. (2022) conducted a large-scale field experiment, finding that embedded training could backfire, and their speculated reason was employees' false sense of security or overconfidence in the company's IT measures.

Our primary objective in this research is not to uncover the theoretical reasons behind the potential impacts of embedded training. Instead, we take an "empirics-first" approach (Golder et al., 2023) and tackle the challenge of "embedded training not as effective as hoped" in its real-world implementations. The "empirics-first" approach originates from a real-world phenomenon or problem, involves the collection and analysis of data, and generates phenomenon-relevant insights without necessarily testing theory. In our context, an unrecognized, inherent shortcoming of embedded training is its limited reach: It reaches only users who fail phishing simulation tests, while users who succeed (but may fail in the future) do not get the training opportunity. To address this practical challenge, we propose non-embedded training as an alternative, novel intervention that companies can adopt in the future. The basic idea of non-embedded training is to decouple post-simulation feedback from landing pages that immediately follow the failure action of user clicking, allowing the feedback to reach *all* users. However, this decoupling leads to delayed timing of the feedback and loses the "just-in-time" benefit of embedded training.[3]

To test the respective and combined effects of embedded and non-embedded training in reducing user vulnerability over time, we used a leading phishing simulation platform and conducted three randomized field experiments over several months across unsuspecting samples of college students. This research contributes to the phishing and cybersecurity

[2] The same "on-the-spot-training" premise is the foundation for other security awareness training interventions that more broadly detect noncompliance behavior and provide immediate coaching; see, for example, KnowBe4's (2025c) Security Coach product and Proofpoint's (2025) ZenGuide product.

[3] An empirics-first approach aligns with a call for phenomenon-focused problematization in IS research (Monteiro et al., 2022) and a need to go beyond the traditional positivist paradigm (Lee, 1991). It can also serve as a stepping-stone to theory development or testing in future research. This approach is particularly well-suited for our investigation based on multiple criteria specified in Golder et al. (2023). First, our research is grounded in the real-world phenomenon of learning by phishing simulations, and it originates from the real-world, "limited reach" problem of embedded training. Second, the literature on the effectiveness of embedded training is equivocal, and we are not aware of any better alternative that has been implemented by practitioners or tested by researchers. Third, each kind of simulation-based training—embedded or non-embedded—has its strengths and limitations, and their respective effectiveness is an empirical question more than a theory-driven question. Finally, our test of an alternative intervention has substantial real-world importance because it provides direct, actionable implications for companies that intend to harness the full power of phishing simulations.

literature by addressing a real-world problem: Given weaker evidence for embedded training in prior field studies, we propose a novel, alternative intervention—non-embedded training—and test how effective this alternative is in reducing user vulnerability. The phishing platform powering our studies is exploring how to incorporate non-embedded training into its next generation of training and coaching tools, attesting to the real-world value and applicability of our research. We also discuss possible theoretical reasons behind our findings, which can be explored in future research. Taken together, our research recognizes both the promise and risk of turning phishing simulations into learning opportunities (Pan et al., 2024; Pienta et al., 2024) and takes an initial step in overcoming the shortcomings of embedded training widely used in the industry.

Related Literature

Phishing Simulations as Training and Embedded Training

Prior IS literature addressing the human side of phishing attacks has investigated both the antecedents of human vulnerability and possible interventions (see the literature review in Section A of the transparency materials).[4] Regarding the latter, some studies have explored various design features such as financial incentives (e.g., Goel et al., 2021) and warning features (e.g., Nguyen et al., 2021). Other studies have examined the effectiveness of different educational programs, such as gamified training systems (e.g., Jensen et al., 2022; Silic & Lowry, 2020), rule-based training, and mindfulness training (e.g., Jensen et al., 2017). Because design features and traditional training programs are outside our scope, we do not delve further into the findings of this stream of studies beyond the exploration of rule-based training.

Rule-based training can be effective for developing user resistance against future attacks if some knowledge about phishing is successfully transferred to the users. To this end, rule-based training involves developing rules of how phishing attempts differ from legitimate emails, and then training users to identify certain cues of phishing attempts and follow the rules (Jensen et al., 2017). Rule-based training can be categorized as list-based if the cues are generic (e.g., "Do not click on a link in an email from an unfamiliar sender") or as example-based if the cues are specific signals that correspond to a particular phishing simulation (e.g., pointing out the unfamiliar sender that users missed during the simulation) (Pan et al., 2024).

Example-based training is closer to our interest—turning phishing simulations into a training opportunity. In Pan et al.'s (2024) investigation, participants were recruited to attend the training session, asked to judge the legitimacy of simulated phishing and legitimate emails, and then were provided with feedback about phishing-related cues that participants may have noticed or missed during the simulation. Their investigation found that example-based training outperforms mindfulness training and list-based training. We consider their example-based training as an implementation of embedded training, but in a highly controlled environment: The users consent first and are conscious of the legitimate vs. phishing classification tasks.

The premise of embedded training is that the most effective feedback is provided at the exact moment when a user fails the simulation and clicks on the phishing link, because "just-in-time" is presumably the most "teachable" moment (Karumbaiah et al., 2016). Not surprisingly, embedded training is widely considered a "best practice" in the cybersecurity industry (Cyberready.com, 2025). As explained by Perry Carpenter, a well-known cybersecurity expert, in his book on transformational security awareness, "getting a baseline of your employees' susceptibility to phishing, and being able to offer just-in-time remediation and training, was groundbreaking for our industry. It was the first true example of implementing just-in-time training at the point of behavior" (Carpenter, 2019, p. 191). Many phishing simulation platforms have built in this function, allowing administrators to customize landing pages in the setup of a phishing simulation campaign (e.g., KnowBe4, 2025b).

Although studies conducted in laboratory settings with scenario-based stimuli have generally confirmed the benefits of embedded training in enhancing users' ability to distinguish phishing from legitimate emails (e.g., Kumaraguru et al., 2007, 2010; Pan et al., 2024), studies conducted in field settings have suggested that these promising benefits may not generalize to real-world implementations (e.g., Ho et al., 2025; Lain et al., 2022). When embedded training is implemented in controlled environments after participants complete simulation exercises (see Pan et al., 2024), *every* participant can get immediate feedback on the phishing-related cues from the exercises. In practice, however, companies often deliver phishing simulation tests without advance notice to obtain more accurate assessments of user susceptibility. By design, the *immediate* feedback in real-world implementations of embedded training cannot be delivered to users who never click on the phishing link (and thus never reach the landing page). Although it is intuitive for companies to ignore users passing a phishing simulation test,

[4] The transparency materials are available at https://osf.io/ut672/

this ignorance represents a lost training opportunity. Prior research shows that user vulnerability is not a stable trait. One study found that nearly half of its employee sample were occasional clickers; employees who didn't click in the first phishing test were still susceptible to doing so in later tests or real phishing attacks (Canham et al., 2021). While this population of users may represent the majority, they would not receive the benefit of embedded training during a phishing simulation. To our best knowledge, no prior research has recognized this limited-reach challenge of embedded training or proposed a better alternative for real-world implementations. We address this challenge by proposing a worthy alternative.

Non-Embedded Training

The limited-reach challenge of embedded training arises from its timing: "just-in-time" feedback delivered at the time of user failure during a phishing simulation. A natural but unrecognized solution would be to decouple feedback on the landing page from the "failure" action of clicking and to deliver the feedback after a delay once the simulation period is over. This alternative intervention, which we call non-embedded training, can address the limited-reach challenge of embedded training because the feedback can be sent to everyone—both users who fail the test and those who pass the test. Thus, non-embedded training increases the proportion of users who get the opportunity to understand their (good or bad) performance in the phishing test and learn relevant knowledge. For users who pass the test, non-embedded training informing them of their performance can reward and reinforce their good behavior (Hattie & Timperley, 2007) because such positive feedback may make them feel good and reduce their likelihood of clicking on future phishing links. Greater exposure to non-embedded training will presumably reduce the vulnerability of all users on aggregate.

Despite this practical advantage, non-embedded training may have its own shortcomings. The primary benefit of embedded training is its "just-in-time" benefit. The time delay of non-embedded training loses this benefit, and it may reduce users' motivation to learn from the failures (or successes) due to the disassociation of the experience from the feedback (Healy et al., 2017). Based on Skinner's reinforcement theory of motivation, the effectiveness of feedback should decline rapidly over time: When the feedback is delayed and decoupled from an event, users tend to be less attached to the event and less motivated to learn (Renner, 1964).

On the other hand, a competing, emotion-based argument would predict an opposite outcome. People have a general tendency to attribute successes to themselves and failures to external factors (Miller & Ross, 1975; Weiner, 2000) because most people want to feel good and see themselves in a positive light (Baumeister, 1998; Sedikides & Strube, 1997). Failures are difficult to manage and threatening to one's self-worth, and threats to the self activate negative emotions (Roese & Olson, 2007). As a result of this emotional barrier in the immediate aftermath of failures, people tend to "tune out" and disengage from the learning opportunity (Eskreis-Winkler & Fishbach, 2019).

As described above, each type of intervention—embedded training and non-embedded training—has its potential strengths and limitations, and even if we disregard their differences in reach, there are competing arguments for the impact of feedback timing. Thus, against a lack of feedback after phishing simulations, the respective and combined effects of embedded and non-embedded training are largely an empirical question. We addressed this practically important question by conducting three randomized field experiments with college students[5] from the business school of a southern U.S. university after obtaining approval from the university's IRB committee. We used an industry-leading phishing platform to send out simulations. Each experiment was longitudinal, occupying an entire semester; the experiments involved an initial phishing simulation followed by embedded and/or non-embedded training and then subsequent phishing simulations to quantify student vulnerability after the intervention.[6] We captured the outcome variable of student vulnerability *multiple times* to take full advantage of the field experimentation method and to understand the decay of intervention effectiveness over time, which has clear implications for practice.[7] Table 1 provides an overview of the studies.

[5] We believe students are adequate because students need to check their email regularly, and they often receive phishing emails (just like company employees). Students are also the participants of choice in many prior field experiments examining the antecedents of phishing vulnerability and ways to reduce it (e.g., Jensen et al., 2017; Wright et al., 2014).

[6] At the time of each experiment, the students had not received institutionally required anti-phishing training or participated in any prior phishing simulations. Importantly, the students were unaware of the ongoing research activity, and we did not ask them to provide informed consent so that we could capture their most natural responses. Meanwhile, following the IRB protocol, we sent a debriefing email to all the students at the end of the experiment to inform them about the experiment and their involvement in it.

[7] We chose different time periods across studies for two reasons: there is no agreed-upon duration of "long-term," and the latest time for delivering the final phishing test is constrained by the length of a semester (e.g., summer semester is shorter). Variation in time periods also extends the external validity of our findings.

Table 1. Overview of Studies

	Purpose	Non-embedded training	Participants	Short-term outcome	Long-term outcome
Study 1	To test the respective effectiveness of embedded and non-embedded training	Conservative implementation (only to failing users)	433 master's students	1 week later	7 weeks later
Study 2	To test the effectiveness of non-embedded training	Aggressive implementation (to all users with reminder)	315 undergrad and master's students	1 week later	4 weeks later
Study 3	To test the respective and combined effectiveness of embedded and non-embedded training	Full implementation (to all users)	2469 undergrad and master's students	4 days later	6 weeks later and 13 weeks later

Study 1

In the first study, we tested the effectiveness of embedded training and a more conservative implementation of non-embedded training—restricting the target audience to only those who failed the phishing test. Our objective was to zoom in to the timing difference between embedded and non-embedded training and to explore whether the just-in-time promise of embedded training can translate into real-world benefits. We randomly assigned 433 master's students to three conditions: control without any feedback, embedded training, and non-embedded training to those who failed the phishing simulation.

Stimulus Materials

We first developed four phishing simulation templates that appear to be different on the surface (see Section B of the transparency materials). We kept the nature and number of actual red flags identical across different simulation templates to achieve equivalence in difficulty levels so that any drop in vulnerability would not be caused by a lower difficulty level. Specifically, each simulation template included five red flags: an email domain spoofing a well-known organization, warnings of negative consequences, a sense of urgency, instructions to click a link, and a link URL not pointing to the actual organization (Moody et al., 2017).

Regarding interventions, the landing page for embedded training and the delayed email for non-embedded training had essentially the same content despite differences in format and timing. The page headline or email subject stated: "You were involved in a simulated [university name] phishing test, and you clicked on it!" The content informed participants that a simulated phishing security test had been conducted, and the participant "did not pass the test and fell for this attack." Then we presented a screenshot of the phishing email revealing the red flags that participants had missed (see Figure 1). The rest of the landing page or email described the relevance and negative consequences of phishing emails and encouraged everyone to "defend our university against cybercrime" (see Section C of the transparency materials for details).

We conducted Study 1 in a spring semester and targeted 433 master's students in their second semester.[8] The average age was 30.4, and 50.1% were male. We loaded the students' names and emails into the phishing delivery platform, and then set up four phishing campaigns to go out at four different dates, as illustrated in Figure 2.

The first phishing test was delivered to everyone to get a basic understanding of each student's vulnerability in preparation for the random assignment of students to different conditions. The test occurred on a Monday early in the semester, and it lasted four days until Friday of that week. If any student clicked on the link contained in the phishing simulation email, they were taken to a seemingly legitimate webpage to enter their username and password. If they entered this sensitive information, the information itself was discarded, and only the behavior was recorded. The final landing page displayed a typical resolution message that did not raise any suspicion. Figure 3 presents the two screenshots of these two webpages for the first phishing test.

[8] Based on a power analysis using G*Power, a sample size of 321 is sufficient to detect a small-to-medium effect size of 0.2 for a three-condition design with 90% power at a significance level of 0.05 (Faul et al., 2007). In this study, we did not target first-semester students to reduce extraneous noise caused by transition into their master's program. Participation rate was 100% because we did not administer informed consent prior to the experiment per the approved IRB application and also because our interventions through landing pages and emails did not require a formal enrollment or "participation" agreement, as would typically be required in a traditional educational program.

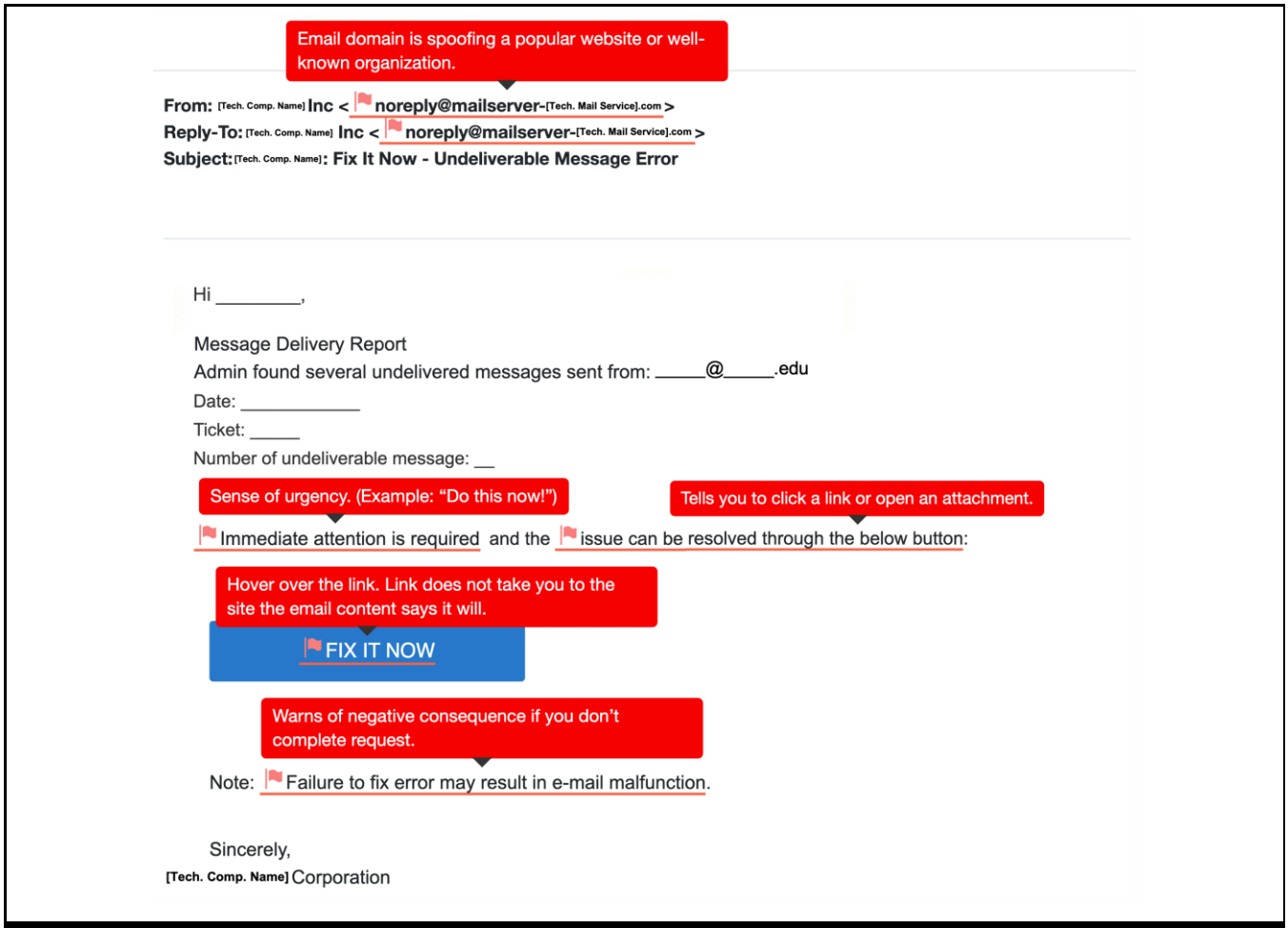


Figure 1. Screenshot of the Phishing Email With Red Flags in Study 1

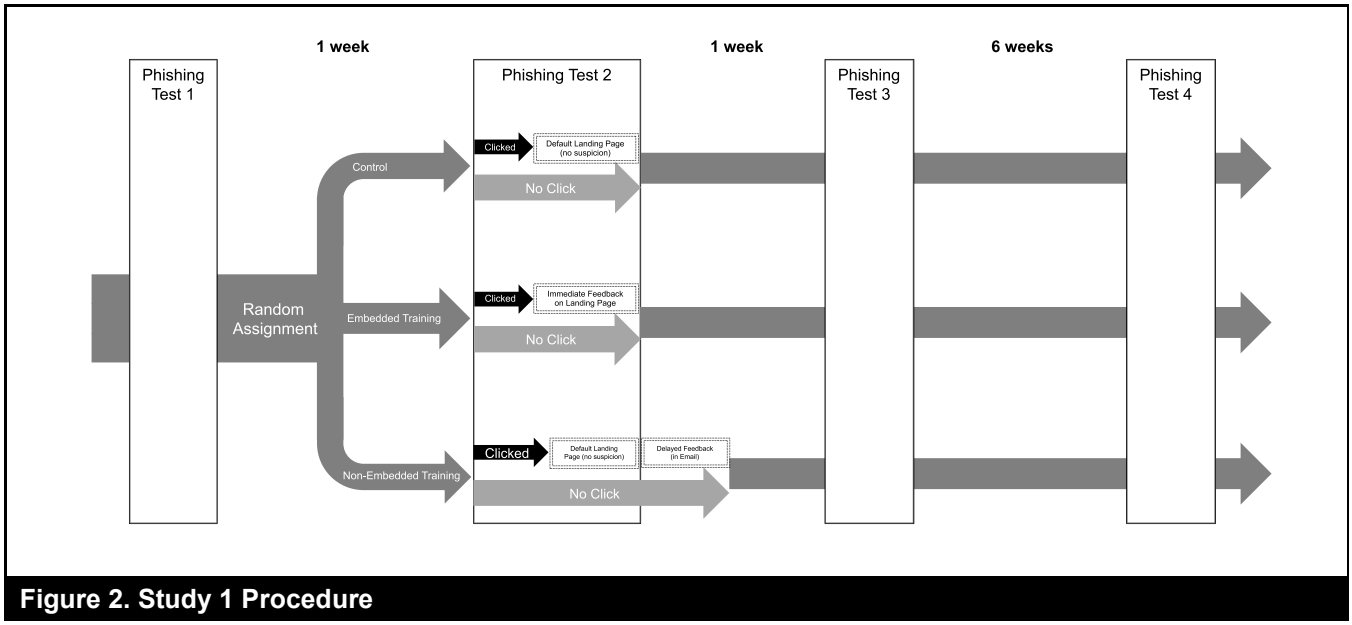
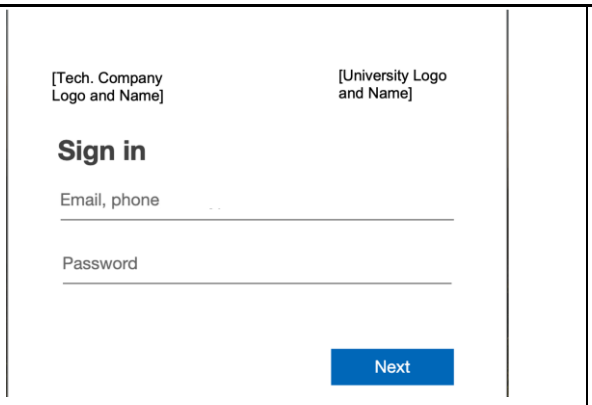


Figure 2. Study 1 Procedure



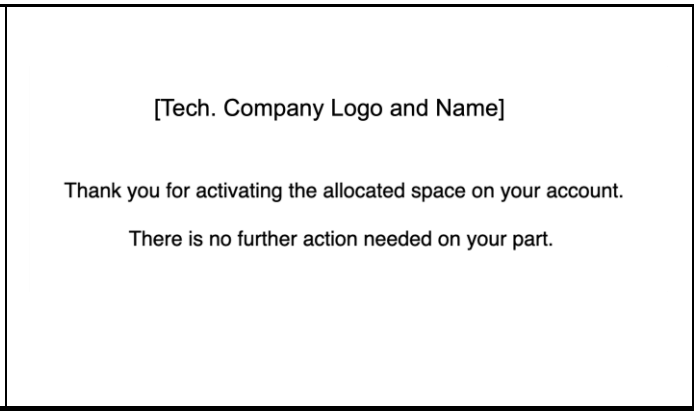


Figure 3. Screenshots of Two Landing Pages After Students Clicked on the Phishing Link

Table 2. Clicked Ratio and Vulnerability (Study 1)

Experimental conditions	Short-term		Long-term	
	Clicked percentage	Vulnerability	Clicked percentage	Vulnerability
Control (*N* = 145)	24.1% (0.429)	0.400 (0.749)	26.2% (0.441)	0.441 (0.781)
Embedded training (*N* = 143)	17.5% (0.381)	0.301 (0.682)	28.0% (0.450)	0.510 (0.846)
Non-embedded training (*N* = 145)	10.3% (0.306)	0.152 (0.476)	17.2% (0.379)	0.303 (0.690)

Note: standard deviations in parentheses.

After the first phishing simulation was over, we randomly assigned each student to one of the three conditions: control, embedded training, and non-embedded training (for failing users only). We conducted block randomization, using each student's department (e.g., information systems, marketing) as a block to ensure that students in each department would be randomly assigned to the three conditions and that their department affiliation could not account for any observable differences. We also verified that students in the three conditions were not significantly different in terms of gender, GPA, or vulnerability in the first phishing test, although age was marginally significantly different between the two treatment conditions.[9]

The second phishing test was delivered about one week after the first test was completed, also on a Monday, and lasted four days. When participants in the control condition clicked on the phishing link, they saw a landing page with a resolution message that did not raise any suspicion. When participants in the embedded training condition clicked on the phishing link, they saw immediate feedback with a red-flag image on the landing page, as described earlier. When participants in the (restricted) non-embedded training condition clicked on the phishing link, they saw the same landing page as those in the control condition and then received an email on the following Monday with the same content as the embedded training.

To quantify intervention effectiveness over time, we conducted the third and fourth phishing tests. We delivered the third phishing test about one week after the second phishing test was over to measure the short-term impact, and we delivered the fourth test about 6 weeks after the third phishing test was over to measure the long-term impact. Both tests were delivered on a Monday, and the landing pages involved two webpages like those in the first phishing test.

Results

To measure the severity of user vulnerability after a phishing test, which is our dependent variable, we coded the variable as 0 if the student never clicked the phishing link, 1 if the student clicked the link but did not enter a username or password, and 2 if the student clicked the link and entered their username or password. We calculated this variable for the last two phishing tests to quantify short-term and long-term impacts. Table 2 presents the percentage of students who clicked the link and means for the vulnerabilities per condition.

[9] See Section D of the transparency materials. Age passed the initial randomization check. After the study was over, however, we noticed the value of zero (indicating missing values) for the birth year of three students. We obtained the birth year and redid the randomization check, finding a marginally significant difference in age between the two treatment conditions. Thus, we controlled for age and all other variables in the data analyses to be more conservative.

Table 3. Study 1 Results

	Short-term vulnerability		Long-term vulnerability	
	Full sample	**Partial sample**	**Full sample**	**Partial sample**
Embedded training	-0.337 (0.315)	-1.709** (0.733)	0.135 (0.270)	-0.629 (0.588)
Non-embedded training (only to failing users)	-1.129*** (0.353)	-2.497*** (0.811)	-0.547* (0.294)	-1.231* (0.638)
Controls	Yes	Yes	Yes	Yes
First cut point	-70.371** (33.030)	-40.908 (73.401)	-8.406 (30.178)	5.486 (63.642)
Second cut point	-69.766** (33.026)	-40.439 (73.392)	-8.035 (30.178)	5.784 (63.642)
N	433	83	433	83
Log likelihood	-217.32	-46.95	-282.42	-63.87
Likelihood ratio $\chi^2(10)$	62.55***	23.04**	24.43***	7.91

Note: Standard errors in parentheses; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The ordered logit model assumes an underlying, unobserved latent variable that determines the outcome, and the cut points are values on this latent variable scale that separate the categories of the outcome. For a given observation, if the value of the latent variable is below the first cut point, the observation belongs to the lowest category (vulnerability = 0). If the latent variable value falls between the first and second cut points, the observation belongs to the second category (vulnerability = 1).

Because the dependent variable is ordinal in nature, we used the ordered logit model:

$$logit\ (\Pr\ (Vulnerability \leq j) = \theta_j - \beta_1 \cdot EmbeddedTraining - \beta_2 \cdot NonEmbeddedTraining - C\gamma,$$

where $Vulnerability$ is the short-term or long-term vulnerability of the students after the intervention and j represents each vulnerability level. $EmbeddedTraining$ is a dummy variable equal to 1 for students in the embedded training treatment condition, and 0 otherwise. $NonEmbeddedTraining$ is a dummy variable equal to 1 for students in the non-embedded training treatment condition, and 0 otherwise. C is a vector of control variables and it includes the following: dummy variables for the student's department, gender, year of birth, GPA, and their vulnerability in the first phishing test. β_1 and β_2 are the parameters of interest.

To apply this equation for each outcome, we ran two separate models: full sample and partial sample. The full sample analysis captures the intention-to-treat (ITT) effect, measuring the mean differences between students assigned to a treatment condition and those assigned to the control condition irrespective of whether the individual clicked on the phishing simulation link during the second test. ITT is practically relevant for organizations because the overall effects of certain interventions are often their primary interest. We also conducted a partial sample analysis, restricting the analysis to only participants who clicked the link in the second test, because only these students had a chance to view the landing page or delayed email that contained our manipulations. Results are presented in Table 3.[10]

Based on the full-sample analyses of ITT, embedded training was ineffective (compared with control) in reducing vulnerability either in the short term or in the long term. A look at partial-sample analyses revealed that embedded training significantly reduced the vulnerability of failing users in the short term (beta = -1.709, $p = 0.020$, odds ratio = 0.181),[11] indicating that it was an effective intervention only for those who clicked on the phishing link, but its narrower reach probably counteracted this benefit and led to a lack of effect in the ITT discussed above. The long-term impact of embedded training using the partial sample remained insignificant.

In contrast, ITT results suggest that (restricted) non-embedded training was effective in reducing vulnerability in the short term (beta = -1.129, $p = 0.001$, odds ratio = 0.323); an odds ratio of 0.323 means that when (restricted) non-embedded training was offered, the odds of student vulnerability in a higher category would decrease by 67.7%, which is a substantial reduction. Moreover, this reduction was still marginally significant in the long term (beta = -0.547, $p = 0.063$, odds ratio = 0.579). The difference in the effects of embedded vs. non-embedded training was also statistically significant in the short term ($\chi^2(1) = 4.58$, $p = 0.032$) and in the long term ($\chi^2(1) = 5.38$, $p = 0.020$). The

[10] We also conducted stepwise analyses including each treatment condition and the control condition to investigate the independent effects of each treatment. The results are qualitatively the same as the results presented in Table 3, so we omitted the stepwise results for brevity.

[11] An odds ratio indicates the odds of an outcome being in a higher category (e.g., from 0 to 1) with a one-unit increase in a predictor variable. An odds ratio of less than 1 indicates that the predictor is associated with lower odds of the outcome being in a higher category.

short-term and long-term impacts of (restricted) non-embedded training using the partial sample were similar to the ITT results, but the difference in the effects of embedded vs. non-embedded training did not reach significance in either term ($p > 0.3$).

Discussion

The first study provided initial evidence that a real-world implementation of embedded training may not be as effective as commonly believed, whereas the delayed feedback in a partial implementation of non-embedded training led to better outcomes.[12] These results challenge the "just-in-time" benefit of embedded training even without considering its limited reach.

Study 2

In the second study conducted during a summer semester, we focused only on non-embedded training and tested its full power delivered to all participants. 315 graduating undergraduate and master's students from the same university who were not involved in the first study participated in this experiment.[13] The average age was 25.8, and 52.7% were male.

This study differed from the first study in the following ways. First, we manipulated one factor: the presence vs. absence of non-embedded training sent to everyone.[14] Second, we administered only three phishing tests in this study, as illustrated in Figure 4. The first phishing simulation was sent to everyone. After the first phishing simulation was over, we randomly assigned participants to one of the two conditions using the same block randomization method as in the first study. We again verified that the participants in the two conditions were not significantly different in (undergraduate or master's) level, gender, birth year, GPA, or vulnerability. After the first phishing test was over and randomization was complete, we sent an intervention email to the participants in the treatment condition, informing them of the phishing test in which they were involved and whether they passed the test ("In particular, a phishing email was sent to you 3 days ago (see screenshot below), and you [passed the test and did not fall / did not pass the test and fell] for this attack."). We also presented a screenshot of the first simulation's red flags (see Section E of the transparency materials for details). In addition, a reminder email with essentially the same information was sent to these participants 6 days later. We added this reminder to strengthen the key manipulation.

Third, we administered the second phishing test to *everyone* one week after the reminder email was sent to the students in the treatment condition. Finally, we administered the third phishing test to everyone three weeks after the second phishing test was over to capture the long-term effects.[15] The landing pages involved a login page followed by a resolution page, just like in the first study. Table 4 presents the clicked ratio of students and mean vulnerabilities.

We again used an ordered logit model in our main analyses, and we used the full sample. We did not conduct additional partial sample analysis as in Study 1 because non-embedded training in this study was sent to *everyone* in the treatment condition. The results are presented in Table 5. Based on the results, non-embedded training (together with a reminder) was effective in reducing participant vulnerability in follow-up phishing tests in both the short term and the long term (beta = -0.775 and -0.770, $p = 0.031$ and 0.048, odds ratio = 0.461 and 0.463).

Note that our manipulation of non-embedded training in the first two studies has at least two limitations. First, our implementation of non-embedded training was conservative in Study 1, targeting only those who failed the test, whereas it was somewhat aggressive in Study 2, which delivered the feedback *twice* to everyone. Second, and most importantly, the designs of the first two studies did not allow us to test the combined effects of the two interventions. We addressed these limitations in the final study by implementing non-embedded training precisely—applying it to everyone without a reminder email and adding a condition that combined both interventions.

[12] An alternative explanation for this effect is that the timing of training provided to participants in the non-embedded training condition (vs. those in the embedded training condition) was closer to the timing of the subsequent (third) phishing test, and this shorter temporal distance might have artificially increased its effectiveness. Yet if this explanation was driving the effect, then the effectiveness of non-embedded training should have been greatly reduced after a long delay; instead, we found it remained effective to some extent in the long term after a substantial delay of nearly two months, indicating its lasting power.

[13] Based on a power analysis using G*Power, a sample size of 266 is sufficient to detect a small-to-medium effect size of 0.2 for a two-condition design with 90% power at a significance level of 0.05 (Faul et al., 2007).

[14] In the actual experiment, we also manipulated another factor independently: the presence (vs. absence) of an opportunity to take some training modules offered by the phishing platform. We decided to collapse different levels of this manipulation because the compliance rate was too low (< 10%), the manipulation of training availability did not have any effect on outcome variables, and it did not interact with feedback manipulation.

[15] We used a shorter delay of 3 weeks in this study (vs. 6 weeks in Study 1) for two reasons: (a) the summer semester was substantially shorter (only 6 weeks), constraining the length of our study, and (b) we could explore the external validity of our findings due to a lack of agreed-upon duration for "long-term" effects.

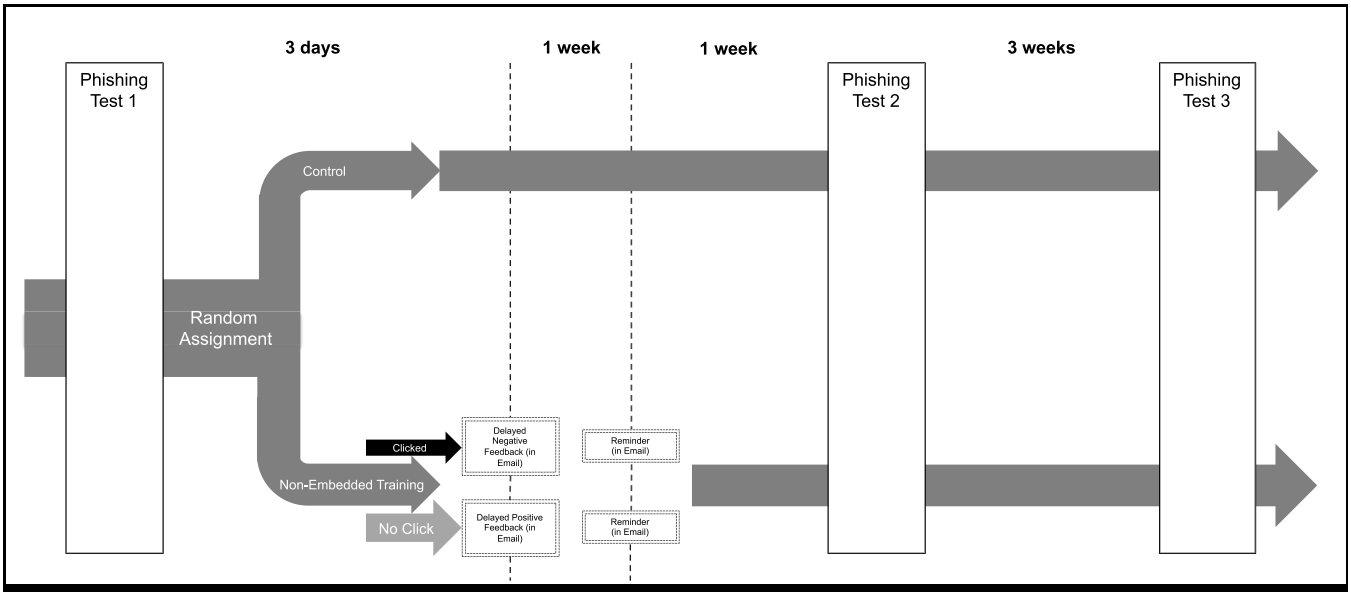


Figure 4. Study 2 Procedure

Table 4. Clicked Ratio and Vulnerability (Study 2)

Experimental conditions	Short-term		Long-term	
	Clicked percentage	Vulnerability	Clicked percentage	Vulnerability
Control (*N* = 156)	17.3% (0.380)	0.276 (0.638)	14.1% (0.349)	0.250 (0.639)
Non-embedded training (*N* = 159)	8.8% (0.284)	0.157 (0.522)	7.5% (0.265)	0.132 (0.479)

Note: standard deviations in parentheses.

Table 5. Study 2 Results

	Short-term vulnerability	Long-term vulnerability
Non-embedded training (with reminder)	-0.775** (0.360)	-0.770** (0.389)
Controls	Yes	Yes
First cut point	2.926 (57.421)	-60.571 (62.486)
Second cut point	3.418 (57.421)	-60.256 (62.484)
N	315	315
Log likelihood	-139.07	-117.12
Likelihood ratio $\chi^2(11)$	18.11*	18.44*

Note: Standard errors in parentheses; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Study 3

In the final study, we targeted 2469 newly enrolled undergraduate and master's students who were not involved in the earlier studies.[16] The average age was 22.3, and 54.9% were male.

This study differed from the earlier studies in the following ways. First, participants in the control condition did not receive any intervention. Participants in the second condition received the embedded training; those in the third condition received non-embedded training; and those in the fourth condition received both interventions (see Figure 5).

[16] Based on a power analysis, a sample size of 1724 is sufficient to detect a small effect size of 0.1 for a four-condition design with 95% power at a significance level of 0.05 (Faul et al., 2007). We used more restrictive criteria (such as an effect size of 0.1 instead of 0.2 and a power of 95% instead of 90%) to test the hypotheses more conservatively. The sample size was originally 2471, but two students opted out after debriefing.

Second, we conducted the random assignment *before* the first phishing test and randomly assigned participants to one of the four conditions using the same block randomization method as in the first study. We again verified that participants in the four conditions were not significantly different in (undergraduate or master's) level, gender, or birth year. We did not use GPA for the random assignment because the participants were newly enrolled, and many did not have a GPA.

Third, we administered the same four phishing tests as in Study 1. The first phishing simulation was sent on a Monday to everyone and lasted four days, and the embedded training was delivered when participants in Conditions 2 and 4 clicked on the link. On Friday of that week when the first phishing test was over, we implemented non-embedded training by sending an email to all the participants in Conditions 3 and 4, informing them of the phishing test, whether they passed the test or not, and red flags they had noticed or missed (see Section F of the transparency materials). We measured short-term vulnerability via the second phishing test sent 4 days after delayed feedback, and long-term vulnerability via the third phishing test sent 5 weeks after the second phishing test was over. To explore the decline of the interventions' effectiveness in the extended long term beyond earlier studies, we also measured student vulnerability via a fourth phishing test sent 7 weeks after the third phishing test was over (see Table 6). The landing pages of these last three tests involved a login page followed by a resolution page, just like in Study 1.

In addition to the full-sample analysis, we also conducted two subsample analyses for those who failed the initial test and those who were successful. The reason is that embedded training could not be delivered to those who failed, whereas non-embedded training could be administered to everyone. Thus, it might be informative to see how each intervention influenced each subgroup of participants differently. The results are presented in Table 7.

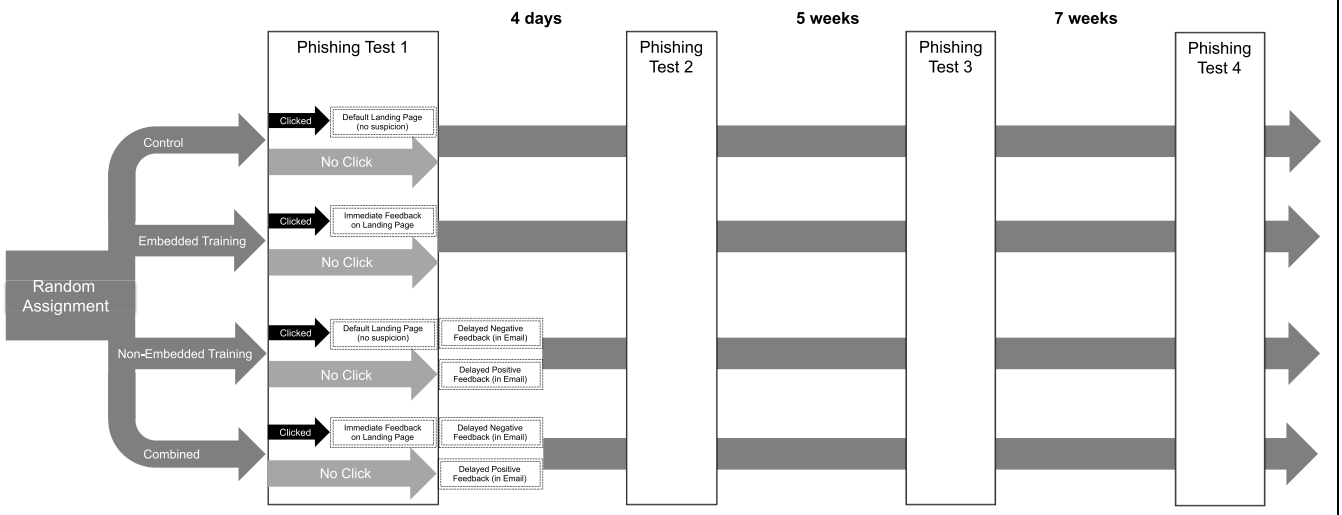


Figure 5. Study 3 Procedure

Table 6. Clicked Ratio and Vulnerability (Study 3)

Experimental conditions	Short-term		Long-term		Extended long-term	
	Clicked percentage	Vulnerability	Clicked percentage	Vulnerability	Clicked percentage	Vulnerability
Control (*N* = 617)	15.4% (0.361)	0.216 (0.541)	10.5% (0.307)	0.154 (0.477)	7.6% (0.265)	0.122 (0.445)
Embedded training (*N* = 618)	8.4% (0.278)	0.117 (0.410)	6.5% (0.246)	0.099 (0.396)	8.3% (0.275)	0.131 (0.460)
Non-embedded training (*N* = 617)	10.2% (0.303)	0.143 (0.451)	7.1% (0.258)	0.102 (0.392)	4.4% (0.205)	0.066 (0.328)
Combined (*N* = 617)	5.5% (0.228)	0.071 (0.314)	6.2% (0.241)	0.091 (0.376)	6.0% (0.238)	0.089 (0.374)

Note: standard deviations in parentheses.

Table 7. Study 3 Results

	Short-term vulnerability			Long-term vulnerability			Extended long-term vulnerability		
	Full sample	Partial sample: failing	Partial sample: succeeding	Full sample	Partial sample: failing	Partial sample: succeeding	Full sample	Partial sample: failing	Partial sample: succeeding
Embedded training	-0.751*** (0.187)	-1.779*** (0.319)	0.048 (0.248)	-0.537** (0.211)	-1.470*** (0.359)	0.164 (0.283)	0.082 (0.213)	-0.686* (0.381)	0.484* (0.268)
Non-embedded training	-0.538*** (0.177)	-0.566** (0.236)	-0.517* (0.284)	-0.446** (0.205)	-0.609** (0.279)	-0.288 (0.314)	-0.642** (0.251)	-0.441 (0.353)	-0.834** (0.371)
Combined	-1.253*** (0.213)	-1.918*** (0.321)	-0.596** (0.291)	-0.614*** (0.214)	-1.805*** (0.386)	0.239 (0.284)	-0.302 (0.230)	-0.253 (0.332)	-0.362 (0.324)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
First cut point	-29.845 (22.858)	-55.612* (33.566)	1.166 (32.324)	-3.206 (29.379)	7.709 (45.266)	-4.900 (38.935)	8.286 (29.629)	18.310 (47.925)	6.187 (38.186)
Second cut point	-28.765 (22.857)	-54.616 (33.562)	2.397 (32.324)	-2.399 (29.379)	8.561 (45.265)	-4.107 (38.935)	8.916 (29.629)	19.013 (47.925)	6.774 (38.186)
N	2469	846	1623	2469	846	1623	2469	846	1623
Log likelihood	-895.98	-407.44	-469.40	-772.33	-310.02	-444.08	-684.96	-274.42	-400.20
Likelihood ratio $\chi^2(11)$	124.85***	84.66***	34.63***	38.49***	49.30***	14.82	48.35***	30.89***	31.81***

Note: Standard errors in parentheses; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Based on the full-sample results of ITT, the intervention of embedded training only (in Condition 2) reduced participant vulnerability in the short term and long term (beta = -0.751 and -0.537, $p < 0.001$ and = 0.011, odds ratio = 0.472 and 0.584), but the effect became insignificant in the extended long term. Subsample analyses indicate that this intervention was effective for those who failed the initial test (as expected), and this effect became weaker and weaker over time (beta = -1.779, -1.470, and -0.686, $p < 0.001$, < 0.001, and = 0.072, odds ratio = 0.169, 0.230, and 0.504).

Regarding the intervention of non-embedded training only (in Condition 3), it reduced participant vulnerability significantly in ITT across all three time periods (beta = -0.538, -0.446, and -0.642, $p = 0.002$, 0.030, and 0.010, odds ratio = 0.584, 0.640, and 0.526). Participants exposed to non-embedded training performed better than those exposed to embedded training in the extended long term ($\chi^2(1) = 8.57$, $p = 0.003$), but not in the short term ($p = 0.289$) or long term ($p = 0.689$). Subsample analyses suggest that the effectiveness of non-embedded training manifested for both failing and succeeding participants in the short term (beta = -0.566 and -0.517, $p = 0.016$ and 0.068, odds ratio = 0.568 and 0.596), for only failing participants in the long term (beta = -0.609, $p = 0.029$, odds ratio = 0.544), and for only succeeding participants in the extended long term (beta = -0.834, $p = 0.024$, odds ratio = 0.434). It appears that both succeeding and failing participants had certain improvements in their performance as time passed by.

Regarding the combined intervention (in Condition 4), ITT results suggest that it was extremely effective in the short term (beta = -1.253, $p < 0.001$, odds ratio = 0.286), as this effect was significantly more effective than either embedded training only ($\chi^2(1) = 4.69$, $p = 0.030$) or non-embedded training only ($\chi^2(1) = 10.17$, $p = 0.001$). However, in the long term, the effectiveness of the combined intervention became weaker (beta = -0.614, $p = 0.004$, odds ratio = 0.541) and was not more effective than either intervention alone ($p > 0.4$). In the extended long term, the effectiveness of combined intervention became insignificant; note that its effectiveness was (marginally) higher than immediate feedback ($\chi^2(1) = 2.89$, $p = 0.09$) but not statistically different from non-embedded training ($p = 0.19$). Subsample analyses revealed that the power of combined intervention (relative to control) in the short term probably came partly from effectively reaching both groups of participants: those who failed (receiving both interventions; beta = -1.918, $p < 0.001$, odds ratio = 0.147) and those who were successful (receiving non-embedded training only; beta = -0.596, $p = 0.040$, odds ratio = 0.551). However, these benefits shrank quickly over time, suggesting that the greater power of combining interventions may not hold over time. The reason for this faster decay over time is worthy of future research.

General Discussion

To explore the real-world effectiveness of embedded training and an alternative intervention—non-embedded training—over time, we conducted three field experiments (see Table 8 for a summary of the ITT results). We found evidence that embedded training may not be as effective as previously assumed (see also Lain et al., 2022), and that non-embedded training is more promising. Moreover, we found in Study 3 that a combination of both interventions may generate the most impact in the short term, but its long-term effectiveness is not as promising as non-embedded training alone.

Table 8. ITT Results Across Studies

Focus	Study 1 (*N* = 433)		Study 2 (*N* = 315)		Study 3 (*N* = 2469)	
	Short-term	Long-term	Short-term	Long-term	Short-term	Long-term
Embedded training	Not effective	Not effective			Effective	Mixed
Non-embedded training	Effective	Effective	Effective	Effective	Effective	Effective
Combined					Effective	Mixed

Practical Implications

Our findings provide direct, actionable insights for both companies and cybersecurity service providers. First, many leading phishing training companies provide training simulations as a way to assess user vulnerability, and such simulations are also used universally in the IS literature to quantify the effectiveness of other interventions (such as mindfulness training; see Jensen et al., 2017). In addition, the landing pages for failing users are often an afterthought or not substantive. We complement Pan et al. (2024) in suggesting that phishing simulations can be turned into a training opportunity in themselves. This switch in focus provides businesses with another opportunity to fight phishing with existing tools, especially when mandatory training is not feasible or when employees are not motivated to start or complete traditional training programs. Given that phishing attacks are a growing threat to firms' cybersecurity and that individual users are the "last line of defense," finding the simplest, lowest-cost approach to enhancing users' anti-phishing resiliency is critical to thwarting attackers. Future phishing research in IS could also consider other motivating effects of simulations beyond assessment.

Second, a best practice for turning phishing simulations into training is embedded training, with the premise that a "just-in-time" approach is the best for teaching users. Further, most phishing simulations completely ignore users who do not "fall" for a given attack. Although a recent study by Lain et al. (2022) suggested that the real-world effectiveness of embedded training may not be as promising, no better alternative has been proposed or tested thus far. We recognize the limited-reach challenge of embedded training and illustrate the potential power of non-embedded training to effectively address this challenge by disassociating the timing of feedback from failing a simulation. Few training companies have considered the benefits of non-embedded training or built such a feature into their simulation platforms. Our research demonstrates the value of this feature, especially for developers of new generations of cybersecurity training, coaching, and feedback tools and services.

Our findings also provide useful guidance for companies on how to best deal with potential phishing risks, given their different priorities. If the priority of a company is the immediate danger of likely phishing attacks, then, based on our evidence, a combined approach (of both embedded training and non-embedded training) may be most effective, but it may need to be repeated to combat the quick decline of its effectiveness. If companies prefer not to repeat phishing simulations too often, which may activate employees' feelings of betrayal (Pienta et al., 2024), then non-embedded training delivered to all employees may have longer-lasting power.

Theoretical Implications

This research also contributes to the growing phishing literature in IS and the broader cybersecurity literature. First, embedded training has been assumed to be ideal, given its intuitive appeal and "just-in-time" nature. Studies in controlled environments have also provided largely consistent evidence supporting its effectiveness in enhancing users' ability to distinguish phishing attempts from legitimate emails (e.g., Kumaraguru et al., 2007, 2010; Pan et al., 2024). Our research challenges this assumption and suggests that embedded training may not be as effective as previously assumed when it is implemented in practice. While Lain et al. (2022) speculated that embedded training may lead employees to feel a false sense of security upon seeing the training information, this explanation cannot fully explain the superior effectiveness of non-embedded training observed in Study 1, where the training message used in both embedded and non-embedded training was identical.

Instead, the implementation of non-embedded training in Study 1 differed from that of embedded training only in feedback timing: immediate or delayed. The discussion that follows centers on the *timing* difference between embedded and non-embedded training, and we use the terms *immediate and delayed feedback* to avoid confusion. With a focus on timing rather than reach, we found in Study 1 that delayed feedback was more effective in reducing user vulnerability than immediate, just-in-time feedback. While this finding is inconsistent with the reinforcement theory of motivation (Skinner, 1954; Skinner & Ferster, 2015), it can perhaps be explained by negative human emotions that activate automatically when people learn about their own failure (Eskreis-Winkler & Fishbach, 2022). Most people dislike failure and tend to attribute failures to external factors to feel good about themselves (Miller & Ross, 1975; Weiner, 2000). In our context, failing users encountering embedded training immediately after clicking on a phishing link

may become emotional and attribute their failures to uncontrollable circumstances that "tricked" them into clicking on the link. Importantly, emotional reactions are crucial in determining how users receive and act upon feedback (Värlander, 2008), and negative emotions can impair the cognitive processing that is needed for learning (Boud & Falchikov, 2007).

In contrast, when the learning opportunity is postponed, as in non-embedded training, the temporal delay serves as a form of psychological distancing, allowing people to move away mentally from what is experienced in the here and now (Kross & Ayduk, 2017). It can help people take a step back from a failure and avoid being overwhelmed by negative emotions, allowing them to reflect on the experience and lessons more adaptively (Kross et al., 2023). After all, emotions are short-lived and dissipate quickly (Ekman, 1992; Frijda, 1993). Moreover, the temporal delay can allow users to "tune out" less and to make sense of their past failure in a more objective manner. In other words, users will likely be more receptive to feedback that teaches them about phishing-related cues that they may have missed, enabling them to learn from the failure.

Attribution and emotion-based theories are not the only candidates for explaining the superior performance of delayed feedback over immediate feedback after a failure, and this superiority may also not be universal. For example, delayed feedback sent in an email might encourage active processing of the feedback at a time of the user's own choosing rather than forcing them to confront it immediately on a landing page. Thus, delayed feedback can enhance users' sense of autonomy (Shute, 2008), which is instrumental for improving learning performance (Dickinson, 1995). Moreover, the superiority of delayed feedback may depend on task difficulty, learner capability, and learning objectives (Mathan & Koedinger, 2002; Schroth, 1992). Red-flag images used in our field experiments may be easy for most students to understand, and the learning goal is the transfer of knowledge rather than mere memorization. Delayed feedback has been found to be beneficial in such situations (e.g., Schroth, 1992). However, when the training material is more difficult, immediate feedback may lead to better performance (e.g., Clariana, 1990). Finally, time can be conceptualized differently, as, for example, clock time, event time, or psychological time, which may influence the impact of feedback timing. Although the exploration of underlying mechanisms, boundary conditions, and alternative forms of time is outside our scope and phenomenon-focused motivation, this is clearly worthy of future research.

Second, to the best of our knowledge, this research is also the first to recognize another inherent limitation of embedded training when it is implemented in practice: limited reach. Although it is intuitive for companies to ignore users who have passed a phishing simulation test and are considered less vulnerable, users who were successful in the past may fail in the future, and the omission of such users may end up hurting companies in the long run. Two of our studies implementing embedded training provided mixed evidence for this limitation. Study 1 found that embedded training can be effective in reducing the vulnerability of failing users in the short term, but it is ineffective overall because users who did not initially fail may fail later. Study 3, based on a larger sample, found that embedded training was effective for all users in the short term, but its long-term effectiveness was mixed. More research is needed to provide more conclusive evidence for this limitation of embedded training.

Third, and more importantly, we propose an alternative intervention—non-embedded training that can be sent to everyone—and tested its efficacy in various forms across three studies. Our results are indeed promising. We found consistent evidence that non-embedded training can be effective in enhancing user resistance due to delayed timing (isolated in Study 1) and greater exposure (in Studies 2 and 3), and its effect remained significant even 3 months after a single intervention (see Study 3). Non-embedded training, by nature, combines delayed timing with greater exposure. Although timing was isolated in Study 1, later studies incorporated both timing and exposure for the implementation of non-embedded training, given our phenomenon-driven focus. Future studies with a phenomenon-driven focus can use design science to develop novel IT artifacts or algorithms (e.g., Abbasi et al., 2021). Future studies with a theory-driven focus could explore one of the two factors in isolation. For example, if timing is not a concern, then studies can fix timing to be delayed and manipulate exposure to reveal more novel insights on the benefits and unintended consequences of expanding exposure under different circumstances.

Future research could also explore more nuanced forms of non-embedded training to generate insights directly relevant for practice. For example, delayed feedback was offered after a few days in our experiments, but what is the optimal amount of delay (e.g., 1 day, a week, or more than a week)? While repeating delayed feedback might enhance the effectiveness of this intervention (as implied in Study 2), is there a threshold at which more repetition backfires? For example, the latest evidence from Pienta et al. (2024) suggests that users may feel betrayed upon learning about simulations, and more frequent simulations may intensify this feeling and introduce unintended consequences. Such explorations will further our understanding of how to best harness the power of simulation-based training opportunities.

Finally, this research complements the pioneering efforts of Pan et al. (2024) and other studies outside IS (e.g., Kumaraguru et al., 2007, 2010) to consider phishing simulations as training opportunities in themselves. Pan et al. (2024) conducted controlled experiments by recruiting participants from online

panels (e.g., MTurk) and implementing example-based training after participants classified a number of phishing and legitimate emails. Kumaraguru and colleagues also conducted such scenario-based experiments in highly controlled settings. However, their findings may be less applicable in real-world situations because people may be more emotional in real life than they would be in an experiment in which they were aware that they were under observation and moving through different scenarios. In addition, the limited-reach shortcoming of embedded training is irrelevant in laboratory studies but is salient in real-world contexts. By implementing embedded training and an alternative intervention in the real world and by capturing their short-term and long-term impacts, this work is among the first to advance our understanding of the practical, longitudinal impacts of such cybersecurity interventions.

Acknowledgements

We sincerely appreciate the senior editor, John D'Arcy, the associate editor, Markus Weinmann, and the anonymous reviewers for their constructive guidance and insightful suggestions. We thank John Just, Xiaoying Xu, Alex Campoe, Bret Masters, and Joann Bower for their generous assistance in the execution of our field experiments. We also thank Stu Sjouwerman and Sidney Fernandes for their support of this research project.

References

Abbasi, A., Dobolyi, D., Vance, A., & Zahedi, F. M. (2021). The phishing funnel model: A design artifact to predict user susceptibility to phishing websites. *Information Systems Research*, *32*(2), 410-436. https://doi.org/10.1287/isre.2020.0973

Abbasi, A., Zahedi, F., Zeng, D., Chen, Y., Chen, H. C., & Nunamaker, J. F. (2015). Enhancing predictive analytics for anti-phishing by exploiting website genre information. *Journal of Management Information Systems*, *31*(4), 109-157. https://doi.org/10.1080/07421222.2014.1001260

Baumeister, R. (1998). The self. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (Vol. 1, pp. 680-740). McGraw-Hill.

Boud, D., & Falchikov, N. (2007). *Rethinking assessment in higher education*. Routledge.

Canham, M., Posey, C., Strickland, D., & Constantino, M. (2021). Phishing for long tails: Examining organizational repeat clickers and protective stewards. *SAGE Open*, *11*(1). https://doi.org/10.1177/2158244021990656

Carpenter, P. (2019). *Transformational security awareness: What neuroscientists, storytellers, and marketers can teach us about driving secure behaviors*. John Wiley & Sons.

Clariana, R. B. (1990). A comparison of answer until correct feedback and knowledge of correct response feedback under two conditions of contextualization. *Journal of Computer-Based Instruction*, *17*(4), 125-129.

Cyberready.com. (2025). *10 best practices for an effective phishing simulation program*. https://cybeready.com/wp-content/uploads/PhishingSimulationsPlaybook.pdf

Dickinson, L. (1995). Autonomy and motivation a literature review. *System*, *23*(2), 165-174. https://doi.org/10.1016/0346-251X(95)00005-5

Ekman, P. (1992). Are there basic emotions. *Psychological Review*, *99*(3), 550-553. https://psycnet.apa.org/doi/10.1037/0033-295X.99.3.550

Eskreis-Winkler, L., & Fishbach, A. (2019). Not learning from failure—the greatest failure of all. *Psychological Science*, *30*(12), 1733-1744. https://doi.org/10.1177/0956797619881133

Eskreis-Winkler, L., & Fishbach, A. (2022). You think failure is hard? So is learning from it. *Perspectives on Psychological Science*, *17*(6), 1511-1524. https://doi.org/10.1177/17456916211059817

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G* power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175-191. https://doi.org/10.3758/BF03193146

Franz, A., Zimmermann, V., Albrecht, G., Hartwig, K., Reuter, C., Benlian, A., & Vogt, J. (2021). Sok: Still plenty of phish in the sea—A taxonomy of user-oriented phishing interventions and avenues for future research. In *Proceedings of the 17th Symposium on Usable Privacy and Security* (pp. 339-358).

Frijda, N. H. (1993). Moods, emotion episodes, and emotions. In L. F. Barrett, M. Lewis, & J. M. Haviland-Jones (Eds.), *Handbook of emotions.* (pp. 381-403). New York, NY, US: Guilford Press.

Goel, S., Williams, K. J., Huang, J. Y., & Warkentin, M. (2021). Can financial incentives help with the struggle for security policy compliance? *Information & Management*, *58*(4), Article 103447. https://doi.org/10.1016/j.im.2021.103447

Golder, P. N., Dekimpe, M. G., An, J. T., van Heerde, H. J., Kim, D. S. U., & Alba, J. W. (2023). Learning from data: An empirics-first approach to relevant knowledge generation. *Journal of Marketing*, *87*(3), 319-336. https://doi.org/10.1177/00222429221129200

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*(1), 81-112. https://doi.org/10.3102/003465430298487

Healy, A. F., Jones, M., Lalchandani, L. A., & Tack, L. A. (2017). Timing of quizzes during learning: Effects on motivation and retention. *Journal of Experimental Psychology: Applied*, *23*(2), Article 128. https://psycnet.apa.org/doi/10.1037/xap0000123

Ho, G., Mirian, A., Luo, E., Tong, K., Lee, E., Liu, L., Longhurst, C. A., Dameff, C., Savage, S., & Voelker, G. M. (2025). Understanding the efficacy of phishing training in practice. In *Proceedings of the IEEE Symposium on Security and Privacy* (pp. 37-54). https://doi.org/10.1109/SP61157.2025.00076

Jensen, M. L., Dinger, M., Wright, R. T., & Thatcher, J. B. (2017). Training to mitigate phishing attacks using mindfulness techniques. *Journal of Management Information Systems*, *34*(2), 597-626. https://doi.org/10.1080/07421222.2017.1334499

Jensen, M. L., Wright, R. T., Durcikova, A., & Karumbaiah, S. (2022). Improving phishing reporting using security gamification. *Journal of Management Information Systems*, *39*(3), 793-823. https://doi.org/10.1080/07421222.2022.2096551

Karumbaiah, S., Wright, R. T., Durcikova, A., & Jensen, M. L. (2016). Phishing training: A preliminary look at the effects of different types of training. In *Proceedings of the 11th Pre-ICIS Workshop on Information Security and Privacy*.

KnowBe4. (2025a). *Best practices: Choosing a landing page*. https://support.knowbe4.com/hc/en-us/articles/115010417488-Best-Practices-Choosing-a-Landing-Page

KnowBe4. (2025b). *Create and manage phishing campaigns*. https://support.knowbe4.com/hc/en-us/articles/204239938-Create-and-Manage-Phishing-Campaigns

KnowBe4. (2025c). *Security coach - real-time coaching in response to risky security behavior*. https://www.knowbe4.com/products/securitycoach

Kross, E., & Ayduk, O. (2017). Self-distancing: Theory, research, and current directions. In J. M. Olson (Ed.), *Advances in experimental social psychology* (Vol. 55, pp. 81-136). Academic Press.

Kross, E., Ong, M., & Ayduk, O. (2023). Self-reflection at work: Why it matters and how to harness its potential and avoid its pitfalls. *Annual Review of Organizational Psychology and Organizational Behavior*, *10*(1), 441-464. https://doi.org/10.1146/annurev-orgpsych-031921-024406

Kumaraguru, P., Rhee, Y., Acquisti, A., Cranor, L. F., Hong, J., & Nunge, E. (2007). Protecting people from phishing: The design and evaluation of an embedded training email system. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 905-914). https://doi.org/10.1145/1240624.1240760

Kumaraguru, P., Sheng, S., Acquisti, A., Cranor, L. F., & Hong, J. (2010). Teaching Johnny not to fall for phish. *ACM Transactions on Internet Technology*, *10*(2), Article 7. https://doi.org/10.1145/1754393.1754396

Lain, D., Kostiainen, K., & Capkun, S. (2022). Phishing in organizations: Findings from a large-scale and long-term study. In *IEEE Symposium on Security and Privacy* (pp. 842-859). https://doi.org/10.1109/SP46214.2022.9833766

Lee, A. S. (1991). Integrating positivist and interpretive approaches to organizational research. *Organization Science*, *2*(4), 342-365. https://doi.org/10.1287/orsc.2.4.342

Mathan, S. A., & Koedinger, K. R. (2002). An empirical assessment of comprehension fostering features in an intelligent tutoring system. In *Proceedings of the International Conference on Intelligent Tutoring Systems* (pp. 330-343). https://doi.org/10.1007/3-540-47987-2_37

Miller, D. T., & Ross, M. (1975). Self-serving biases in the attribution of causality: Fact or fiction? *Psychological Bulletin*, *82*(2), 213-225. https://doi.org/10.1037/h0076486

Monteiro, E., Constantinides, P., Scott, S., & Shaikh, M. (2022). Qualitative research methods in information systems: A call for phenomenon-focused problematization. *MIS Quarterly*, *46*(4), i-xviii.

Moody, G. D., Galletta, D. F., & Dunn, B. K. (2017). Which phish get caught? An exploratory study of individuals' susceptibility to phishing. *European Journal of Information Systems*, *26*(6), 564-584. https://doi.org/10.1057/s41303-017-0058-x

Nguyen, C., Jensen, M. L., Durcikova, A., & Wright, R. T. (2021). A comparison of features in a crowdsourced phishing warning system. *Information Systems Journal*, *31*(3), 473-513. https://doi.org/10.1111/isj.12318

Pan, S., Kwak, D.-H., Kuem, J., & Kim, S. (2024). Roles of feedback and phishing characteristics in antiphishing training performance: Perspectives of goal-setting and skill acquisition. *Journal of the Association for Information Systems*, *25*(4), 1037-1078. https://doi.org/10.17705/1jais.00854

Pienta, D., Thatcher, J. B., Wright, R. T., & Roth, P. L. (2024). An empirical investigation of the unintended consequences of vulnerability assessments leading to betrayal. *Journal of the Association for Information Systems*, *25*(4), 1079-1116. https://doi.org/10.17705/1jais.00875

Proofpoint. (2025). *Simplify behavior change with automated, risk-based learning*. https://www.proofpoint.com/us/products/mitigate-human-risk

Renner, K. E. (1964). Delay of reinforcement: A historical review. *Psychological Bulletin*, *61*(5), 341-361. https://doi.org/10.1037/h0048335

Roese, N. J., & Olson, J. M. (2007). Better, stronger, faster: Self-serving judgment, affect regulation, and the optimal vigilance hypothesis. *Perspectives on Psychological Science*, *2*(2), 124-141. https://doi.org/10.1111/j.1745-6916.2007.00033.x

Schroth, M. L. (1992). The effects of delay of feedback on a delayed concept formation transfer task. *Contemporary Educational Psychology*, *17*(1), 78-82. https://doi.org/10.1016/0361-476X(92)90048-4

Sedikides, C., & Strube, M. J. (1997). Self-evaluation: To thine own self be good, to thine own self be sure, to thine own self be true, and to thine own self be better. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 29, pp. 209-269). Academic Press.

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, *78*(1), 153-189. https://doi.org/10.3102/0034654307313795

Silic, M., & Lowry, P. B. (2020). Using design-science based gamification to improve organizational security training and compliance. *Journal of Management Information Systems*, *37*(1), 129-161. https://doi.org/10.1080/07421222.2019.1705512

Skinner, B. F. (1954). The science of learning and the art of teaching. *Harvard Educational Review*, *24*, 86-97.

Skinner, B. F., & Ferster, C. B. (2015). *Schedules of reinforcement*. BF Skinner Foundation.

Värlander, S. (2008). The role of students' emotions in formal feedback situations. *Teaching in Higher Education*, *13*(2), 145-156. https://doi.org/10.1080/13562510801923195

Vijayalakshmi, M., Mercy Shalinie, S., Yang, M. H., & U, R. M. (2020). Web phishing detection techniques: A survey on the state-of-the-art, taxonomy and future directions. *IET Networks*, *9*(5), 235-246. https://doi.org/10.1049/iet-net.2020.0078

Weiner, B. (2000). Intrapersonal and interpersonal theories of motivation from an attributional perspective. *Educational Psychology Review*, *12*, 1-14. https://doi.org/10.1023/A:1009017532121

Wright, R. T., Jensen, M. L., Thatcher, J. B., Dinger, M., & Marett, K. (2014). Influence techniques in phishing attacks: An examination of vulnerability and resistance. *Information Systems Research*, *25*(2), 385-400. https://doi.org/10.1287/isre.2014.0522

Author Biographies

Dezhi Yin is an associate professor of information systems and Muma Fellow at the Muma College of Business, University of South Florida. He received his Ph.D. in information technology management from the Georgia Institute of Technology. His research interests include emotions in online environments, user-generated content (e.g., online word-of-mouth), and AI-generated content (e.g., chatbots). He has published in *MIS Quarterly, Information Systems Research, Journal of Management Information Systems, Journal of the Association for Information Systems, Journal of Marketing Research, Academy of Management Journal, Production and Operations Management*, and other academic journals. He is a recipient of the 2024 AIS Mid-Career Award and the 2021 INFORMS ISS Sandra A. Slaughter Early Career Award. He serves as an associate editor for *MIS Quarterly*, and he received the 2023 Outstanding Associate Editor Award and the 2017 Reviewer of the Year Award from the same journal.

Matthew Mullarkey is a professor of instruction, information systems, in the Muma College of Business at the University of South Florida. He received his Ph.D. in information systems from the University of South Florida, an M.B.A. from the University of South Carolina, an M.S. in systems management from the University of Southern California, and a B.S. in engineering from the United States Military Academy. His research is centered on the guided, emergent design of innovative technology systems, processes, products, and services across disciplines and industries, including healthcare, manufacturing, technology, defense, energy, non-profit, and government. He has published in *MIS Quarterly*, *INFORMS Journal on Computing*, *European Journal of Information Systems*, *IEEE Transactions on Engineering Management*, *Information Systems Frontiers*, and *Communications of the Association for Information Systems*, among others. He is a Fulbright Core Research Scholar, an Extraordinary Professor in the Centre for Applied Research in Management at the Northwest University (South Africa), and an IIIS Interdisciplinary Fellow. He is the co-founder and inaugural director of the USF Doctor of Business Administration degree program, and the co-founder and inaugural executive director of the USF-TGH People Development Institute. Mullarkey has been awarded more than $7 million as the principal investigator (PI) in industry partnership grants and $4.85 million as PI or Co-PI on competitive NSF and DoD grant-funded research.

GJ de Vreede is a professor of information systems and dean of the School of Business at Stevens Institute of Technology. Previously, he served as associate dean and interim dean at the Muma College of Business of the University of South Florida and as director of the Center for Collaboration Science at the University of Nebraska at Omaha. He received his Ph.D. in information systems from Delft University of Technology, the Netherlands. His research focuses on AI in teams, crowdsourcing, and Collaboration Engineering. His research has appeared in journals such as *Information Systems Research, Journal of Management Information Systems, Journal of the Association for Information Systems, Information & Management, Management Information Systems Quarterly Executive, Small Group Research,* and *Communications of the ACM*. He currently serves as editor-in-chief of the *Group Decision & Negotiation* journal.

Moez Limayem is president of the University of South Florida. His research focuses on information systems, technology adoption, and the transformative role of digital innovation in organizations and society. He has published extensively in leading journals, including *MIS Quarterly*, *Information Systems Research*, and *Journal of Management Information Systems*. Dr. Limayem has also served in senior academic leadership roles at major research universities, where he has advanced research excellence, student success, and global engagement.