JOURNAL OF MANAGEMENT INFORMATION SYSTEMS
2022, VOL. 39, NO. 3, 793–823
https://doi.org/10.1080/07421222.2022.2096551

Routledge
Taylor & Francis Group

# Improving Phishing Reporting Using Security Gamification

Matthew L. Jensen[a], Ryan T. Wright[b], Alexandra Durcikova[a], and Shamya Karumbaiah[c]

[a]Price College of Business, University of Oklahoma, Norman, OK, USA; [b]McIntire School of Commerce, University of Virginia, Charlottesville, VA, USA; [c]School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, USA

**ABSTRACT**
Phishing is an increasing threat that causes billions in losses and damage to productivity, trade secrets, and reputations each year. This work explores how security gamification techniques can improve phishing reporting. We contextualized the cognitive evaluation theory (CET) as a kernel theory and constructed a prototype phishing reporting system. With three experiments in a simulated work setting, we tested gamification elements of validation, attribution, incentives, and public presentation for improvements in experiential (e.g., motivation) and instrumental outcomes (e.g., hits and false positives) in phishing reporting. Our findings suggest public attribution with rewards and punishments best balance the competing necessities of accuracy with widespread reporting. Furthermore, our results demonstrate the unique benefits of security gamification to phishing reporting over and above other phishing mitigation techniques (e.g., training and warnings). However, we also noted that unintended consequences in false alarms might arise from shifts in motivation resulting from public display of incentives. These findings suggest that carefully calibrated external incentives (rather than intrinsic rewards) are most likely to improve the ancillary task of phishing reporting.

**KEYWORDS**
Phishing; leaderboards; cognitive evaluation theory; online gamification; security gamification; phishing reporting; incentives

## Introduction

Among the most prominent threats to organizational information security is phishing [42], where attackers masquerade as legitimate message senders and try to steal credentials and private information. In 2020, the U.S. Federal Bureau of Investigation (FBI) reported a 69 percent increase in Internet crime, including $1.8 billion in losses due to phishing attacks [31]. Phishing attacks also enabled criminals to encrypt and hold for ransom data at large companies [e.g., 53] and municipalities [e.g.,11, 12]. The impact of these phishing attacks can be dire. For example, the city of New Orleans declared a state of emergency as losses from cyberattacks totaled more than $7 million [85].

To counter phishing, scholars have developed and organization leaders have invested in automated tools (e.g., filters, machine learning detection) and training and awareness initiatives [17, 46]. Although such investments have provided essential gains, phishing attacks *still* penetrate organizational defenses. Therefore, organizations have implemented

**CONTACT** Matthew L. Jensen mjensen@ou.edu Price College of Business, University of Oklahoma, 307 West Brooks, Norman, OK 73019-4006, USA.

Supplemental data for this article can be accessed at www.tandfonline.com/mmis.
Supplemental data for this article can be accessed online at https://doi.org/10.1080/07421222.2022.2096551.

systems that facilitate real-time employee reporting of phishing attacks. These phishing reports help inform machine learning models and other detection tools while also providing threat intelligence to information technology (IT) security teams [18]. Phishing reporting also increases employees' joint responsibility for IT security, which contributes to a vigilant, resilient organization that quickly detects, mitigates, and recovers from known and emergent cyber threats [22, 23]. However, phishing reporting is often an extra-role behavior that falls outside of formal work responsibilities and information security policies (ISPs) and, therefore, may be viewed as voluntary [39, 40, 41]. Furthermore, enlisting and motivating organization members has proved difficult because reporting cybersecurity threats is often tricky, inconvenient, and requires skills that some employees lack [58, 81]. If scholars and practitioners are to realize the potential benefits that come from real-time reporting of phishing messages, they must learn how to motivate this extra-role, protective behavior.

In a novel modification to phishing training, gamification has been proposed as a viable mechanism to motivate organization members and improve their phishing detection [73, 74]. *Gamification* uses game design elements in non-game contexts [28]. Gamification has also been shown to improve the performance of IT security activities aside from phishing training [29]. In parallel, gamified reporting systems are beginning to be marketed by IT security companies (e.g., SANS Institute, Proofpoint, and Barracuda) as a mechanism to enlist users and motivate phishing reporting and other protective behaviors [64]. These combined efforts have introduced the concept of *security gamification* as a method of enlisting organization members in IT security efforts. Although a myriad of commercial security applications using gamification have emerged, a considerable amount of variation in capabilities and features exists in these systems (see Online Supplemental Appendix 1) and researchers have revealed mixed results when investigating the efficacy of gamification initiatives [7]. As a result, researchers have called for a systematic investigation of gamification elements intended to motivate and improve secure behavior [52, 53, 73]. Therefore, we focus on how organization leaders can use security gamification, a fundamental design decision [51], to facilitate the reporting of phishing messages in a way that builds *motivation* and reporting accuracy (measured by *hits* and *false positives*).

To address our research purpose, we developed a prototype gamified system that uses a leaderboard to facilitate reporting of phishing attacks. The prototype relies on Cognitive Evaluation Theory [CET; 24, 26] as a kernel theory. CET is well suited in this context because it describes the effects of feedback, external rewards, and competition on motivation and performance. We contextualize CET, following Hong et al. [38], to understand how different types of feedback affect motivation and performance. This theoretical lens allows us to focus our attention on 1) the feedback structure of gamification elements provided as part of a leaderboard (e.g., attribution and validation of reports); 2) incentives for reporting (e.g., external rewards and punishments); and 3) presentation of gamification elements that may induce competition (e.g., public or private). Finally, we examined the effect of a leaderboard among other common anti-phishing interventions (e.g., training and awareness campaigns) to demonstrate proof of concept.

We devised a pilot experiment to test theoretical mechanisms and three randomized experiments to understand resulting motivation and performance changes. In the experiments, participants received legitimate and phishing messages in a simulated work setting. During the first two experiments, we manipulated the feedback structure and presentation of gamification elements and incentives. Findings from the first two experiments informed the last experiment during which we examined the efficacy of a leaderboard alongside three formal training and awareness techniques widely used to mitigate phishing attacks: 1) video-based training, 2) marking emails as "external" when they come from outside the organization, and 3) marking emails with warnings if there is a high probability they are phishing. In sum, these three experiments illuminate the potential benefits of security gamification using a leaderboard and CET's application as a kernel theory. However, we also reveal critical tradeoffs that come with using gamification to motivate and improve phishing reporting.

## Related Work

Even though the benefits of manual detection of phishing attacks have been debated [e.g., 15, 16], both detection and reporting of phishing attacks are critical to an organization's security. Manual detection and reporting counter new permutations of attacks that may slip through automated defenses, inform actions and policies the organization may take to prevent future attacks, and serve as inputs for improvements in automated defenses. Reporting phishing attacks is prototypical of ancillary, extra-role activities to which individuals often devote insufficient attention and are viewed as voluntary [39, 61]. Thus, reporting phishing attacks may often be supplanted by other higher priority work tasks. Even when individuals report phishing attacks, lack of feedback on one's own and others' performance may reduce opportunities to learn from past correct and incorrect reports, thus short-circuiting potential future improvement [5]. Therefore, the process and outcomes of phishing reporting in the workplace are often viewed as a black box [84].

Practitioners and researchers have developed several techniques to assist in the manual detection of phishing attacks. Among the most popular are digital training, warnings about external message senders, and warnings about suspected phishing [74]. Digital training is intended to reduce successful attacks by building competence in detecting phishing messages [43]. Warnings about external senders and suspected phishing messages are designed to increase awareness and scrutiny of potential phishing messages that otherwise would likely be processed as trustworthy email messages [60, 84]. The introduction of these interventions has improved phishing detection, and these interventions are now commonplace recommendations for organizational best practices and standards [62]. However, scholars have noted limitations of training and awareness campaigns and attacks still slip through [4].

Furthermore, although training and awareness interventions build competence, they do little to motivate accurate phishing reporting and provide no feedback regarding the accuracy of reports. Without reporting, organizations are deprived of critical inputs that help protect against existing and emerging attacks. Therefore, we investigate a technique designed explicitly to motivate mundane activities and give individuals feedback: gamification.

### Gamification, Motivation, and Task Performance

*Gamification* is traditionally used to describe the use of game design elements in non-game contexts [28]. Its purpose is to motivate the performance of a task and enhance individuals' experience while performing the task [33]. For example, standard gamification techniques include providing feedback via points and leaderboards to structure and incentivize the performance of a task [e.g., 6]. As participants engage with the focal task, they build competence, deepen engagement, and increase motivation to progress in the game. Studies have shown gamified systems increase task engagement [7, 68], elevate attitude toward a task [77], improve individuals' efficacy and autonomy, and increase positive outcomes for organizations [6, 67, 70]. Gamification has been studied in extra-role behaviors such as incentivizing zero-waste within the workplace with a great deal of success [21]. For example, Oppong-Tawiah et al [65] found that gamified apps reduced electricity use in a traditional office setting. While Gartner predicted the broad adoption of gamified systems within organizations [13], it also indicated that 80 percent of these gamified systems would fail due to poor design elements [34].

Research has also suggested that motivation and performance gains resulting from gamified systems may be more nuanced than previously thought [7, 37]. Interestingly, some organizations experienced difficulties implementing gamification. Some gamified systems did not produce their intended benefits [35, 78] and decreased motivation when attempting to incentivize corporate responsibility initiatives [55, 78]. Furthermore, incentives and explicit comparisons of individuals on leaderboards were shown to build external motivation but came at the cost of intrinsic motivation and building competence [57]. Gaming scholars have even argued that users of gamified systems are sometimes exploited, coerced, or manipulated [e.g., 9, 10].

More recently, scholars in IT security have used gamification techniques to improve phishing protection within organizations [29, 73]. In one example, using a six-month field study, researchers demonstrated that a gamified training system increased motivation, learning, and efficacy against attacks [73]. This work has led to the concept of *security gamification* or using "game-like design artifacts and system processes to strengthen employees' motivations to encourage learning, efficacy, and increased employee compliance with organizational security initiatives" [73, p. 131]. Although initial findings for security gamification are promising, prior research has thus far focused on phishing training, not on the actual reporting of phishing messages, which confers organization-wide protection. Moreover, " . . . more research needs to examine each element involved in gamified security training. We studied an entire system, but each part should also receive further attention." [73, p. 153]. Successful gamified systems strictly align gamification design elements (e.g., game objects and mechanics) and target system components (e.g., user/task) to achieve desirable user-system interactions [51]. User-system interactions then promote positive experiential outcomes (e.g., motivation) and instrumental outcomes (e.g., task performance) [51]. The effect of security gamification elements and how to align them with system components is currently unknown yet is critical when developing phishing reporting systems [73]. Therefore, we investigate the connections between security gamification elements, the system components, and outcomes.

### Kernel Theory: Cognitive Evaluation Theory

To disentangle the effects of elements in security gamification, we turn to CET, which serves as a kernel theory for our investigation. It describes the primary antecedents of intrinsic and external motivation and explores the effects of feedback, external rewards, and competitions on motivation and task performance [see 26 for a review]. Thus, CET addresses several mechanisms understood to determine the effects of gamification elements [51]. CET predicts that intrinsic motivation leads to superior outcomes and is based on people's need to be competent and self-determining while engaging in an optimally challenging task [24, 25]. When individuals are intrinsically motivated, task performance is self-sustaining, competence increases, and task performance improves [26]. However, when competence or autonomy is threatened or when the task is too challenging (or not challenging enough), intrinsic motivation is undermined, and task performance may need to be incentivized. Specifically, CET predicts that negative feedback and performance-contingent incentives can threaten needs for competence and autonomy [23, 79]. However, negative feedback and incentives are sometimes necessary when building competence, so CET predicts that they will not damage intrinsic motivation so long as they are administered and perceived informationally [26]. If intrinsic motivation is somehow damaged, external motivation for building competence will be required, and locus of causality for action will shift toward the source of the external motivation (e.g., incentives, competition) [26]. Although dramatic in its effect on action [69], external motivation may also result in unintended consequences as task performance becomes contingent on external factors and is no longer self-sustaining.

To maximize the utility of gamified systems, the gamification literature calls for strict alignment between context and gamification elements [e.g., 51]. In answer, we followed the guidelines presented by Hong et al. [38] on contextualizing general theory. Contextualization promotes the application and customization of general theories to a specific phenomenon of interest (see Table 1). Such application illuminates the boundaries and assumptions of the

**Table 1.** Development of contextualized theory.

| Guideline Steps | Explanation | Examples in Security Gamification |
|---|---|---|
| 1: Identify and ground in a general theory | " . . . built on a general theory that is applicable to the research domain of interest" [38, p. 117] | CET offers antecedents to motivation and predictions for outcomes based on motivation. |
| 2: Contextualize and refine a general theory | "The general theory needs to be contextualized to the specific domain . . . " including developing a minimal set of "core constructs relevant to a particular context" [38, p. 117] | Core constructs: Feedback, external rewards, competition.<br>Contextualizing details: Security is an ancillary, extra-role behavior that may need to be incentivized. Phishing reporting often lacks feedback. Process and outcomes of phishing reporting are viewed as a black box. Security gamification is beneficial for training. |
| 3: Evaluate the context-specific factors | Context-specific factors should be identified and vetted | Gamification elements were developed based on CET constructs: validation, attribution, rewards and punishments, public presentation. Theoretical effects of gamification elements determined by a pilot study. |
| 4: Model context-specific factors | Context-specific factors are developed into a theoretical model | Hypotheses were developed to provide theoretical relationships between the gamification elements decomposed from CET and the context-based DVs. |
| 5: Examine the interplay between the IT artifact and other factors | "Interactions among context-specific factors pertaining to the specific technology [. . .] should be examined". [38, p. 117] | Three experiments were designed and executed to test these contextualized relationships and related interactions. |

*Note*: Adapted from Hong et al. [38]. Hong et al. [38] call for an additional step 6 to examine alternative context-specific models. This is necessary "when the objective is to examine the indirect factors" [p. 117], which is beyond the scope of our work.

Table 2. Definition and measurement of key variables.

| Theoretical Concepts | Variables | Definition |
|---|---|---|
| **Gamification Elements** | | |
| Feedback (CET) | Validation | Feedback from an authoritative source informing individuals if they were correct or incorrect to report a message |
| External Rewards (CET) | Attribution | Disclosure of a contributor's identity who reported a message |
| | Reward | Points given for correctly identifying a phishing message. |
| | Punishment | Points taken for incorrectly identifying a phishing message. |
| Competitions (CET) | Public Presentation | Public display of gamification elements (e.g., validation, attribution, rewards, and punishments) for a group to observe |
| **Outcomes** | | |
| Experiential | Motivation | An individual's desire to find and report phishing messages (Subjective) |
| Instrumental | Hits | Correctly reported phishing messages (Objective) |
| | False Positives | Incorrectly reported phishing messages (Objective) |

general theory and results in a greater understanding of the phenomenon. We first identified the core constructs of CET (e.g., feedback, external rewards, and competition) and matched gamification elements with each construct (see Table 2). Next, we refined CET and incorporated context specific factors unique to phishing reporting. Finally, we formalized hypotheses based on our contextualized theory and tested these predictions.

## Hypotheses

After someone reports a message as suspicious, *validation* is the response individuals receive from an authoritative source informing them if they were correct or incorrect to report the email message. Pioneers in gamification [e.g., 56] have argued that validation is essential for gamified systems because it provides the feedback mechanism to evaluate participants' learning and performance. However, CET predicts that feedback may have conflicting effects on intrinsic motivation and subsequent task performance [26]. On the one hand, feedback limits autonomy and shifts an internal locus of causality to be more external, shifting motivation to also be more external. On the other hand, feedback builds competence and increases intrinsic motivation during the performance of an optimally challenging task. This tradeoff is particularly salient for gamified phishing mitigation systems because individuals will likely receive negative feedback as they report messages that appear suspicious but are in fact legitimate.

CET resolves these conflicting effects by suggesting that if the feedback is delivered and perceived to be informational, the negative effects of control are overcome by the positive effects of building competence [26]. On its own, validation in a gamified system provides feedback in an informational way that minimizes control. Using validation, individuals are notified concerning the correctness of their actions, have correct actions reinforced, and have the chance to rectify incorrect actions in the future. Thus, informational feedback in validation helps build competence that will increase motivation and improve performance. In addition, knowing which reports were accurate will assist individuals as they decide on and strengthen their own decision criteria for identifying phishing messages. This strengthening of individuals' decision criteria should increase the number of hits and reduce false positives.

*Hypothesis 1 (H1): Validation will increase (a) motivation to report phishing attacks and (b) hits and will decrease (c) false positives.*

The second element is the *attribution* of a report in the gamified system to a single participant. When a participant correctly or incorrectly reports a suspicious message, attribution ties the contributor's identity to the report. Attribution is an extrinsic reward offered by the organization [40] and can be used to build reputation and as evidence of expertise [20, 82]. In gamified systems, attribution can switch the motivation from intrinsic to external as the assignment of an individual's identity to each report ties performance to ego. Thus, attribution is a controlling mechanism that is predicted to damage intrinsic motivation but significantly elevate external motivation [69]. As attribution shifts the locus of causality toward the source of the contingent external reward, individuals will likely attempt to garner as many hits as possible in pursuit of reputational gains. However, CET is ambiguous concerning external motivation's effect on false positives. Research on deception detection in other digital contexts (e.g., human resource systems) suggests that false positives rise when individuals are prompted to increase hits [8]. In their pursuit of external rewards, individuals may lower their threshold of suspicion for what constitutes a reportable message. In other words, the goal of contingent rewards may increase hits but come at the expense of increased false positives. We expect similar effects to occur with gamified phishing reporting systems.

*Hypothesis 2 (H2): Attribution will increase (a) motivation to report phishing attacks, (b) hits, and (c) false positives.*

Organization leaders can use incentives to promote certain behaviors along with gamification elements, including validation and attribution. Separate from validation (which describes the correctness of the behavior), *incentives* describe rewards and punishments that individuals receive if they perform a correct (or incorrect) behavior. Organizations have long used direct, tangible incentives (e.g., financial bonuses) to promote desired behavior [49, 50, 89]. However, gamified systems frequently include intangible incentives (e.g., gaining points) that recognize expertise and reputation [30]. Such performance-contingent rewards and punishments reduce autonomy and shift the locus of causality toward the source of the contingency and size of the reward [26]. With gamified systems where rewards and punishments are doled out in intangible points, strong external motivation is predicted to replace intrinsic motivation. Thus, as individuals pursue points and the reputational benefits points represent, the number of hits is expected to rise. However, as with the pursuit of other external rewards (e.g., attribution), the pursuit of points is likely to result in increased false positives along with hits.

Intangible punishments (e.g., losing points) are expected to have the opposite effect from rewards. As punishments are enacted for the incorrect performance of the desired behavior, individuals will be more likely to view the performance of the desired behavior as a mandate with punishments allotted for any deviations. Compulsion in a gamified system will likely stifle motivation and undermine attempts of the desired behavior. Individuals will be unlikely to report phishing messages unless the message surpasses a high threshold of suspicion. This would result in fewer hits but would also result in fewer false positives.

*Hypothesis 3 (H3): Rewards (punishments) will increase (decrease) (a) motivation to report phishing attacks, (b) hits, and (c) false positives.*

Finally, *public presentation* is a display of gamification elements such that all participants in the gamified system can view them. While private display of validation, rewards, and punishments provides valuable feedback to facilitate learning and update criteria for distinguishing phishing, public display provides the same feedback but also makes performance relative to others highly salient. Although both presentation types depict a controlling mechanism (i.e., feedback), significant differences can affect motivation and performance. Private presentation of gamification elements is more likely to be interpreted as contributing to competence, thus overcoming the controlling nature of feedback. Feedback presented in private is expected to be perceived as informational rather than autonomy-reducing and will likely foster intrinsic motivation. Private feedback also sets up an indirect competition in which individuals compete with themselves to pursue greater task competence [26]. But, when gamification elements are public, the competition is direct and unambiguously against peers. Public presentation shifts motivation from intrinsic to external and facilitates social comparisons of performance. Public presentation is also likely to increase accountability through identifiability and awareness of logging [80]. Public logging of performance initiates social comparisons and accountability, which are potent sources of external motivation [26, 32] and are likely to increase hits as participants seek favorable positions relative to peers. As with attribution, we predict that individuals will lower their threshold of suspicion for reportable messages in response to public presentation. Thus, external motivation brought on by social comparison will result in increased false positives.

*Hypothesis 4 (H4): Public presentation of gamification elements will increase (a) motivation to report phishing attacks, (b) hits, and (c) false positives more than private presentation.*

According to CET, other anti-phishing efforts such as training and phishing awareness interventions should improve phishing reporting by increasing competence. For example, training instructs individuals on distinguishing between malicious and legitimate messages and warnings about suspicious email messages or messages originating outside of the organization alert individuals and provide diagnostic information that could improve detection and subsequent reporting. Additionally, training and awareness interventions preserve autonomy and direct attention to suspected phishing messages. However, these interventions do not directly promote or increase motivation to report phishing messages.

In other maintenance tasks (e.g., preventative health), mere knowledge of beneficial behaviors is often insufficient in precipitating action; motivation must also be present if action is to occur [59]. A similar observation has also been made in security gamification: gamified systems motivate action in ways that text-based training does not [73]. Like other forms of phishing mitigation (e.g., a priori training and warnings), leaderboards can support learning and increase awareness of potential phishing attacks. As predicted by CET, gamified systems also offer a structured approach for reporting and receiving feedback on reported messages. But training, which typically includes feedback on practice detection tasks, and warnings, which provide no feedback, do not. Furthermore, different from other types of phishing mitigation, leaderboards offer a *reason* to report suspected messages, thus inducing participation.

Consistent with contextualized CET and our aforementioned theorizing, a gamified system that implements a leaderboard using public attribution with rewards and punishments builds motivation by connecting to explicit, external rewards (e.g., reputational gains). Therefore, the leaderboard will likely generate significant motivation to report phishing attacks that is over and above what is introduced by other phishing mitigation techniques. Public attribution with rewards and punishments will also combine to shift the locus of causality toward the source of contingent external rewards. Therefore, individuals will be likely to report more hits than would be reported with other phishing mitigation techniques. But as individuals pursue external rewards, they risk reporting a greater number of legitimate messages as phishing.

> *Hypothesis 5 (H5): Use of a leaderboard implementing public attribution with rewards and punishments will increase (a) motivation to report phishing attacks, (b) hits, and (c) false positives beyond effects of other phishing mitigation techniques (digital training, external warnings, and phishing warnings).*

## Design Model

We conducted a pilot experiment and three controlled, randomized laboratory experiments to test our hypotheses and vet our design of a gamified phishing reporting system (total N = 711). Although several methods could have been used to investigate the elements of gamified systems, lab experiments offer high internal validity and the ability to systematically examine interventions that serve as the basis for information systems [27]. In addition, lab experiments are often necessary for establishing proof of concept before field studies and adoption in field settings can occur [63]. Finally, experimental research shows causal linkages that confirm (or disconfirm) theory [50], a critical need in advancing security gamification research [73].

The pilot experiment (N = 143) investigated the motivations (e.g., intrinsic, external) and autonomy people felt when they viewed leaderboards manipulated in Experiments 1-3. Results indicated that consistent with CET and our arguments, intrinsic motivation was dominant from validation, external motivation was dominant from attribution, both intrinsic and external motivation dropped when punishments were introduced alongside rewards, and external motivation was dominant from public presentation (see Online Supplemental Appendix 2). However, we did not observe the damaging effect to intrinsic motivation and autonomy that CET predicts should occur in response to attribution and public presentation. Instead, intrinsic motivation increased in response to attribution and public presentation. Therefore, based on the results of this pilot experiment, the manipulations were deemed to be successful. Additionally, since intrinsic and external motivation responded in the same direction to manipulations, we captured general motivation in the main experiments.

Experiment 1 (N = 104) investigated gamification elements of validation and attribution via a 2 (validation vs. no validation) x 2 (attribution vs. no attribution) factorial experiment. Experiment 2 (N = 171) investigated incentives and presentation of feedback to participants via a 3 (rewards only vs. punishments only vs. both rewards and punishments) x 2 (public vs. private presentation) factorial experiment. Experiment 3 (N = 293) compared the performance of the leaderboard to other phishing mitigation techniques in a 2 (leaderboard vs. no leaderboard) x 2 (training vs. no training) x 2 (external warning vs. no external warning) x 2 (phishing warning vs. no phishing warning) factorial experiment. See Online Supplemental Appendix 5 for an a priori power analysis.

## Experiment Procedures

All three experiments shared the same procedure and had two phases. First, participants were invited to be part of what they were told was an email usage study. After agreeing to participate, the first phase was a pre-survey that contained covariate items and asked participants to schedule a time to come to the lab. The second phase was a lab session during which participants were asked to assume the role of an intern to a senior vice president (SVP) of a software company called PrepDesign and manage the SVP's email inbox. Approximately one week after the pre-survey, participants arrived at the lab, were consented, oriented to their tasks, and given a list of employees and personal contacts for the SVP. Participants responded to messages from other executives, scheduled meetings for the SVP, and forwarded personal emails to the SVP's account. Next, participants were instructed to help plan a future product marketing event by finding three different hotels in a remote city that could handle the event. The messages in each inbox and all work tasks were the same for all participants. These work tasks were meant to simulate multiple organizational priorities that employees must manage in actual work settings.

At the beginning of the lab session, participants were randomly assigned to condition and given an intern number used during the experiment (e.g., Intern001). An experimenter using a script introduced the task and Outlook Web Access and then allowed participants to practice within the environment. As part of their orientation, participants reviewed information security policies specifying that suspected phishing messages should be reported and describing how to report them. When participants began the experiment and opened their inboxes, eight emails were waiting to be processed, one of which was a phishing message. Eighteen additional emails were sent to participants during the session, including four additional phishing messages. Phishing emails were modeled after real malicious messages [48]. They mimicked an IT-service desk request, a cloud storage share request, a deal from a hotel chain, a payment receipt, and a security alert (see Online Supplemental Appendix 3 for legitimate and phishing messages included in the experiments). All phishing emails contained links to web pages we owned. If participants clicked on a link in a phishing email, they were first directed to one of our web pages (where their computers could have been compromised if the phishing attack were real). Then they were immediately redirected to a legitimate website. Participants had a total of 30 minutes to process all 26 messages and complete the work tasks (including event planning), after which they were directed to a post-survey. Finally, participants were instructed not to share details of the experiment with others and were dismissed. To ensure consistency across experiments, we developed custom software to manage the distribution and reporting of emails and observe participants' inboxes and calendar activities. Experiment procedures were piloted twice with nonoverlapping student samples to refine the instructions, number, content of messages, experiment manipulations, and work tasks.

## Outcome Measures

The same dependent variables and covariates were used in all experiments to draw conclusions across experiments. The first dependent variable was experiential and captured *motivation* to report phishing messages using several survey items (see Online Supplemental Appendix 4). Next, we captured the accuracy of phishing reports using two instrumental measures. When participants suspected a message they received was phishing, they were instructed to report it by forwarding it to the IT security department. Similar to past phishing and deception research [14, 47], we measured accuracy in terms of *hits* (correctly reported actual phishing messages) and *false positives* (legitimate messages reported as phishing messages).

Past research has demonstrated that factors aside from our manipulated variables also influence individuals' identification of phishing messages [e.g., 86, 87]. Therefore, we captured propensity to trust [66], perceived Internet risk [44, 54], internal and external computer self-efficacy [76], and self-reported expertise in identifying phishing messages, as these variables have been used in recent phishing research [e.g., 45].

We assessed the measurement properties of the scales in Experiments 1-3 following recommendations from Brown [12]. All scales demonstrated satisfactory convergent and discriminant validity. Across the three experiments, motivation violated assumptions of normality; therefore, we applied a $(k + x)^3$ transformation for parametric analysis. The complete results of the measurement models and the individual items are presented in Online Supplemental Appendix 4. Additionally, we performed robustness checks with our treatments to ensure that leaderboard conditions did not reduce work productivity or increase perceived work disruption. These analyses are reported in Online Supplemental Appendix 6.

## Experiment 1: Validation and Attribution

To test H1 and H2, students from business classes at a large U.S. mid-western university were recruited for an experiment and were offered course credit for their participation. Although previous researchers have expressed concerns over using student subjects [19], our sample was appropriate for three reasons. First, students are a suitable representation of young working professionals. Many are currently in the workforce, and nearly all will shortly join the workforce. Second, all of them used email during school or work and were familiar with its function. Finally, students are a common target of phishing attacks [75].

A total of 120 students completed phase 1; however, 16 students did not attend phase 2 and were excluded from the study. Therefore, a total of 104 participants completed the experiment. They reported a mean age of 20.7 (SD = 2.2), a mean of 2.5 (SD = .7) years of education after high school, and, of all participants, 27.9 percent were female. Indicating their suitability as participants, those in Experiment 1 had experience with phishing attacks. For example, 33.7 percent of participants reported knowing someone who had been phished, and 34.6 percent reported nearly falling for phishing themselves.

### Experiment 1: Independent Variables

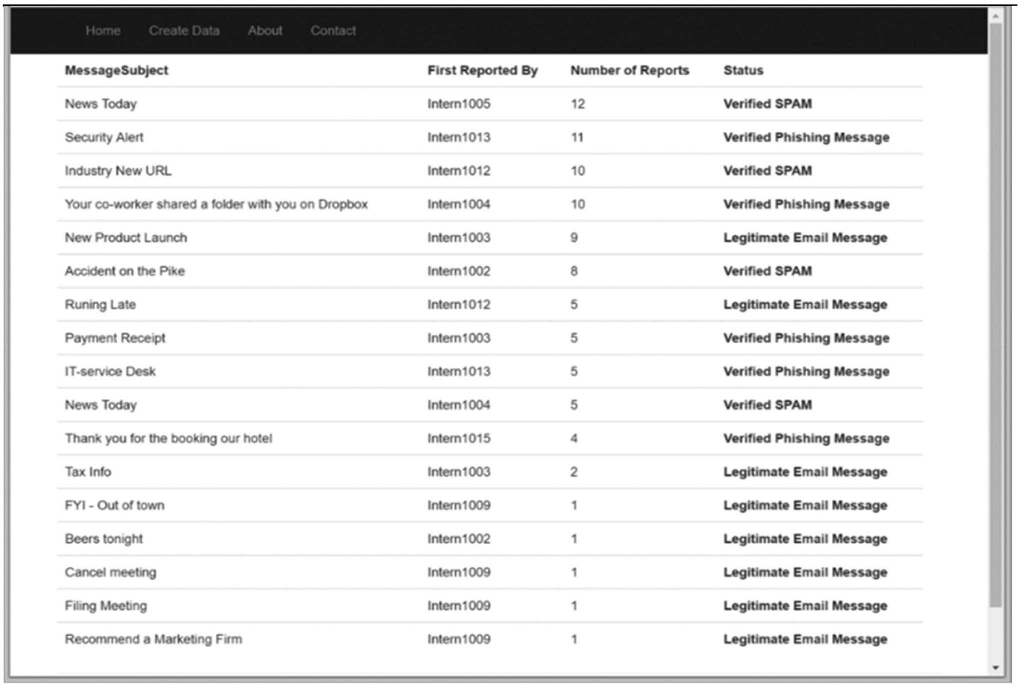
Experiment 1 was a 2 (validation: validation vs. no validation) x 2 (attribution: attribution vs. no attribution) factorial design. When a message was reported as possible phishing, the message's subject line was displayed on the leaderboard. As more individuals reported the same message as phishing, the number of reports was also displayed. Those in the validation condition were shown additional information concerning whether reported messages were actually phishing. For 90 seconds after their initial report, messages were labeled "Under Review." Then the status was updated to either "Verified Phishing Message," "Confirmed

SPAM," or "Legitimate Email Message." Those in the attribution condition were shown the intern number of the participant who first reported the message. An example message board with validation and attribution elements is shown in Figure 1. In conditions without attribution and validation, the "First Reported By" and "Status" columns were hidden (respectively).

## *Experiment 1: Evaluation*

Table 3 shows the means for each dependent variable in each condition for Experiment 1. A total of 98 participants (94.2 percent) reported at least one suspicious message, and the mean number of unique messages reported per person was M = 5.38, SD = 2.83.

To test H1 and H2, we performed a multiple analysis of covariance (MANCOVA) with validation and attribution as the independent variables and propensity to trust, perceived internet risk, computer self-efficacy internal, computer self-efficacy external, and phishing experience as covariates. The complete results of the MANCOVA for Experiment 1 are in Online Supplemental Appendix 5.



| MessageSubject | First Reported By | Number of Reports | Status |
|---|---|---|---|
| News Today | Intern1005 | 12 | Verified SPAM |
| Security Alert | Intern1013 | 11 | Verified Phishing Message |
| Industry New URL | Intern1012 | 10 | Verified SPAM |
| Your co-worker shared a folder with you on Dropbox | Intern1004 | 10 | Verified Phishing Message |
| New Product Launch | Intern1003 | 9 | Legitimate Email Message |
| Accident on the Pike | Intern1002 | 8 | Verified SPAM |
| Runing Late | Intern1012 | 5 | Legitimate Email Message |
| Payment Receipt | Intern1003 | 5 | Verified Phishing Message |
| IT-service Desk | Intern1013 | 5 | Verified Phishing Message |
| News Today | Intern1004 | 5 | Verified SPAM |
| Thank you for the booking our hotel | Intern1015 | 4 | Verified Phishing Message |
| Tax Info | Intern1003 | 2 | Legitimate Email Message |
| FYI - Out of town | Intern1009 | 1 | Legitimate Email Message |
| Beers tonight | Intern1002 | 1 | Legitimate Email Message |
| Cancel meeting | Intern1009 | 1 | Legitimate Email Message |
| Filing Meeting | Intern1009 | 1 | Legitimate Email Message |
| Recommend a Marketing Firm | Intern1009 | 1 | Legitimate Email Message |

**Figure 1.** Sample leaderboard for Experiment 1.

**Table 3.** Means for the dependent variables in Experiment 1.

| Condition | | N | Motivation (SD) | Hits (SD) | False Positives (SD) |
|---|---|---|---|---|---|
| No Attribution | No Validation | 25 | -.21 (.85) | 3.48 (1.12) | 2.12 (1.94) |
| | Validation | 25 | -.26 (1.06) | 2.68 (1.89) | 2.04 (2.13) |
| Attribution | No Validation | 30 | .15 (.60) | 2.53 (1.53) | 2.30 (1.71) |
| | Validation | 24 | .30 (.41) | 3.54 (1.41) | 2.96 (1.52) |

Multivariate tests indicated significant effects for attribution, $F(3,93) = 3.966$, $p = .010$, $\eta_p^2 = .11$, and the attribution x validation interaction approached significance, $F(3,93) = 2.647$, $p = .054$, $\eta_p^2 = .08$. Univariate tests were then used to interpret the multivariate effects. They revealed a significant medium effect from attribution on motivation, $F(1,95) = 8.456$, $p = .005$, $\eta_p^2 = .08$ and a significant attribution x feedback interaction with a medium effect on hits, $F(1,95) = 7.989$, $p = .006$, $\eta_p^2 = .08$. To interpret significant univariate tests, we conducted a series of pairwise comparisons with a Bonferroni correction. The significant effects for attribution are provided in Table 4 and the significant attribution x feedback effect on hits is shown in Figure 2.

Since validation did not have a significant main effect on motivation (H1a), hits (H1b), or false positives (H1c), these hypotheses were not supported. However, validation (along with attribution) was involved in an interaction that impacted the number of hits. Therefore, the conditions producing the most hits had both validation and attribution, or neither was included.

In support of H2a, attribution increased motivation to contribute to the gamified system. However, no main effect on hits (H2b) was observed. Instead, attribution was involved in a significant interaction influencing hits, as previously noted. Attribution increased the number of hits, but only when validation was present. Attribution did not demonstrate significant effects on false positives (failed support for H2c). Finally, supplemental analysis (see Online Supplemental Appendix 6) did not uncover negative effects from the leaderboard (e.g., on work productivity and disruption).

### *Experiment 2: Rewards, Punishments, and Public Presentation*

Participants were recruited from introductory business classes at a large U.S. mid-western university and were offered course credit for their participation. A total of 274 people began phase 1, but 103 participants did not complete the pre-survey or did not attend phase 2 and were excluded from the study. This resulted in 171 participants. Analysis of excluded participants indicated no threat from non-response bias (see Online Supplemental Appendix 6). Included participants reported a mean age of 20.9 (SD = 3.8), a mean of 2.3 (SD = .8) years of college education, and 65.5 percent were male. Like Experiment 1, participants in Experiment 2 were frequently exposed to phishing attacks. For example, 35.7 percent of participants reported knowing someone who had fallen for a phishing message, and 33.9 percent of participants reported nearly falling for a phishing message themselves.

**Table 4.** Experiment 1 post-hoc comparisons of attribution conditions on motivation, hits, and false positives.

| Dependent Variable | Condition | Comparison | Mean Difference (Condition – Comparison) | Standard Error | Significance |
|---|---|---|---|---|---|
| Motivation[a] | Attribution | No Attribution | 25.882 | 8.900 | .005 |
| Hits | Attribution | No Attribution | -0.095 | .301 | .752 |
| False Positives | Attribution | No Attribution | 0.569 | .375 | .132 |

*Note*: [a]Reflects transformed values for motivation.

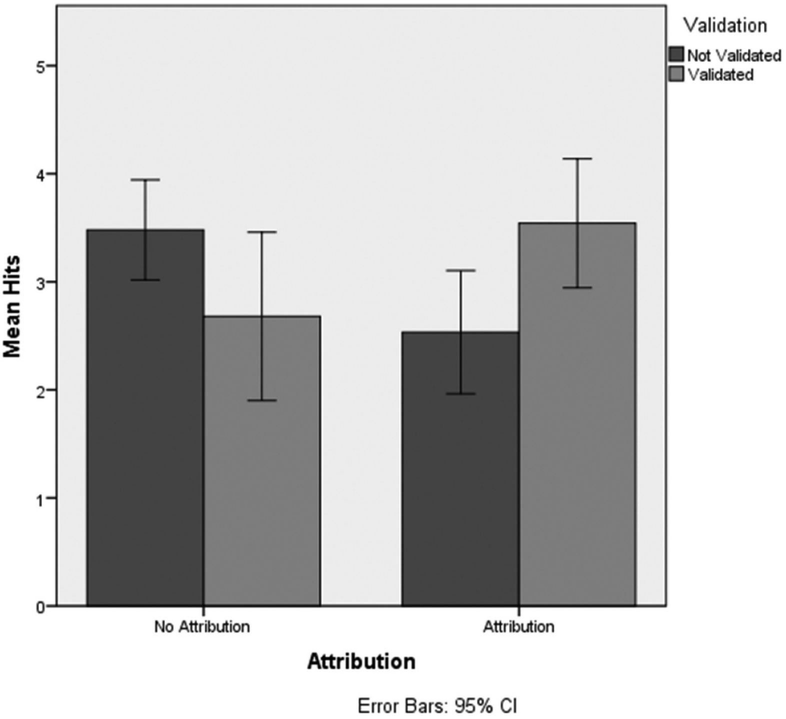


**Figure 2.** Experiment 1: Interaction of attribution and validation on number of hits.

### *Experiment 2: Independent Variables*

H3 and H4 were tested in Experiment 2 using a 3 (incentive: rewards only vs. punishments only vs. both rewards and punishments) x 2 (presentation: public vs. private) factorial design. As incentives for correctly reporting phishing messages, participants were awarded points. Intangible points are a common element in gamified systems [56, 83] and, beyond validation addressed in Experiment 1, directly quantify the correct performance of the desired action compared to incorrect or non-performance. Participants in the rewards only condition who correctly reported a phishing message (a hit) were awarded 100 points on the leaderboard, and no deductions were made. Participants in the punishments only condition who incorrectly reported a phishing message (false positive) were deducted 25 points on the leaderboard, and no additions were made.[i] Participants in both rewards and punishments conditions received points for correct reports and lost points for incorrect reports. The imbalance between the reward for hits and punishment for false positives indicated the greater relative importance of hits over false positives. Total points for hits and false positives were determined based on piloting and consultation with industry partners.

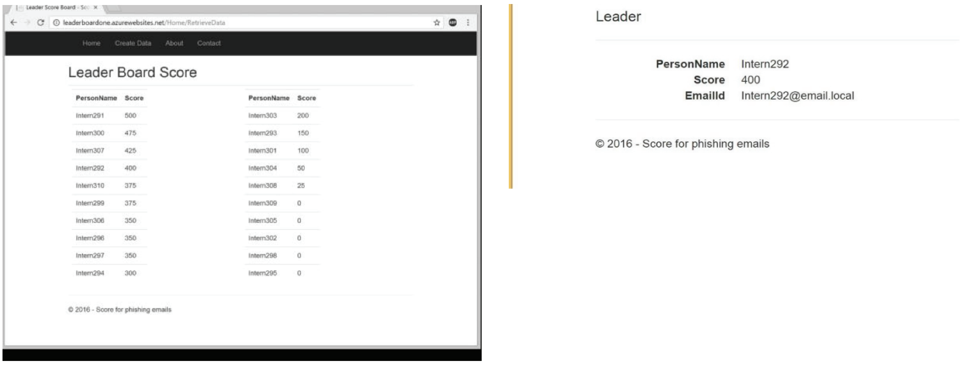
Participants in the public condition had their score publicly posted for all participants in the experiment session to see, and participant scores were sorted on the leaderboard, highest to lowest. Each participant in the private condition was shown only his or her own performance on an individualized webpage. Since attribution was shown to be effective in elevating motivation and the number of hits, participants' intern numbers were displayed next to their scores in all conditions. The two conditions of presentation are shown in Figure 3.

### Experiment 2: Evaluation

Table 5 shows the means for each dependent variable in each condition for Experiment 2. A total of 134 participants (78.4 percent) reported at least one suspicious message, and the mean number of unique messages reported per person was M = 3.94, SD = 2.66.

To test for differences in motivation, hits, and false positives (H3-H4), we performed a MANCOVA with incentives and presentation as the independent variables and propensity to trust, perceived internet risk, internal computer self-efficacy, external computer self-efficacy, and phishing experience as covariates. The complete results of the MANCOVA analysis for Experiment 2 are reported in Online Supplemental Appendix 5.

Multivariate tests indicated significant results for presentation, $F(3,158) = 2.778$, $p = .043$, $\eta_p^2 = .05$, incentives, $F(6,318) = 3.999$, $p = .001$, $\eta_p^2 = .07$, and the incentive x presentation interaction, $F(6,318) = 2.625$, $p = .017$, $\eta_p^2 = .05$. Next univariate tests were used to interpret the multivariate effects. They revealed a significant medium effect from incentives on motivation, $F(2,160) = 6.177$, $p = .003$, $\eta_p^2 = .07$, hits, $F(2,160) = 4.717$, $p = .010$, $\eta_p^2 = .06$, and false positives, $F(2,160) = 8.821$, $p < .001$, $\eta_p^2 = .10$. A significant small effect from presentation was observed on hits, $F(1,160) = 4.646$, $p = .033$, $\eta_p^2 = .03$, and false positives, $F(1,160) = 6.339$, $p = .013$, $\eta_p^2 = .04$. Finally, a significant medium incentive x presentation effect was observed on hits, $F(2,160) = 5.229$, $p = .006$, $\eta_p^2 = .06$.



Leader Board Score

| PersonName | Score |
|---|---|
| Intern291 | 500 |
| Intern300 | 475 |
| Intern307 | 425 |
| Intern292 | 400 |
| Intern310 | 375 |
| Intern299 | 375 |
| Intern306 | 350 |
| Intern296 | 350 |
| Intern297 | 350 |
| Intern294 | 300 |

| PersonName | Score |
|---|---|
| Intern303 | 200 |
| Intern293 | 150 |
| Intern301 | 100 |
| Intern304 | 50 |
| Intern308 | 25 |
| Intern309 | 0 |
| Intern305 | 0 |
| Intern302 | 0 |
| Intern298 | 0 |
| Intern295 | 0 |

© 2016 - Score for phishing emails

Leader

| | |
|---|---|
| PersonName | Intern292 |
| Score | 400 |
| EmailId | Intern292@email.local |

© 2016 - Score for phishing emails

Public

Private

**Figure 3.** Presentation conditions for Experiment 2.

**Table 5.** Means for the dependent variables in Experiment 2.

| Condition | | N | Motivation | Hits | False Positives |
|---|---|---|---|---|---|
| Private | Reward Only | 27 | .18 (.76) | 2.07 (1.62) | 1.96 (1.51) |
| | Punishment Only | 27 | .01 (.87) | 1.93 (1.44) | 1.26 (1.35) |
| | Both Reward and Punishment | 25 | -.22 (.84) | 1.32 (1.31) | 1.32 (1.03) |
| Public | Reward Only | 31 | .35 (.62) | 2.94 (1.84) | 2.81 (1.35) |
| | Punishment Only | 25 | -.46 (.98) | 1.36 (1.35) | 1.64 (1.25) |
| | Both Reward and Punishment | 36 | .02 (.87) | 2.69 (1.51) | 1.72 (1.28) |

To interpret significant univariate tests, we conducted a series of pairwise comparisons with a Bonferroni correction. The results of the pairwise comparisons for incentives are summarized in Table 6. The results of the pairwise comparisons for presentation are summarized in Table 7. The significant incentive x presentation interaction effect on hits is plotted in Figure 4.

Consistent with our theorizing for H3a and H3b, the reward-only condition produced higher motivation and more hits than the punishment only condition. But the condition containing both rewards and punishments was no better than the punishment only condition for both motivation and hits. The results show that the introduction of punishments (even when joined with rewards) dampens motivation and depresses the number of hits. However, as shown in Figure 4, the significant incentive x presentation interaction revealed nuance for the effect on hits. Consistent with our theorizing, those in the public rewards and punishment condition produced more hits than those in the private rewards and punishment condition and those in the punishment only condition. This pattern of results partially supports H3a and H3b.

In a similar vein, participants in the punishment only and rewards and punishments conditions produced fewer false positives than those in the reward only condition. Again, this finding was consistent with H3c. However, no differences were noted in the number of false positives between the punishment only condition and rewards and punishments condition. Therefore, H3c was partially supported.

**Table 6.** Experiment 2 post-hoc comparisons of incentive conditions on motivation, hits, and false positives.

| Dependent Variable | Condition | Comparison | Mean Difference (Condition – Comparison) | Standard Error | Significance |
|---|---|---|---|---|---|
| Motivation[a] | Reward Only | Punishment Only | 32.877 | 10.044 | .004 |
| | | Both Reward, Punishments | 26.200 | 9.726 | 0.023 |
| | Punishment Only | Both Reward, Punishments | -6.677 | 10.085 | 1.000 |
| Hits | Reward Only | Punishment Only | 0.881 | .289 | 0.008 |
| | | Both Reward, Punishments | 0.503 | .280 | 0.223 |
| | Punishment Only | Both Reward, Punishments | -0.378 | .290 | 0.583 |
| False Positives | Reward Only | Punishment Only | 0.943 | .254 | 0.001 |
| | | Both Reward, Punishments | 0.862 | .246 | 0.002 |
| | Punishment Only | Both Reward, Punishments | -0.081 | .255 | 1.000 |

*Note*: [a]Reflects transformed values for motivation.

**Table 7.** TT7>Experiment 2 post-hoc comparisons of presentation conditions on motivation, hits, and false positives.

| Dependent Variable | Condition | Comparison | Mean Difference (Condition – Comparison) | Standard Error | Significance |
|---|---|---|---|---|---|
| Motivation[1] | Public | Private | -.767 | 8.203 | .926 |
| Hits | Public | Private | .509 | 0.236 | .033 |
| False Positives | Public | Private | . 522 | 0.207 | .013 |

*Note*: Reflects transformed values for motivation.

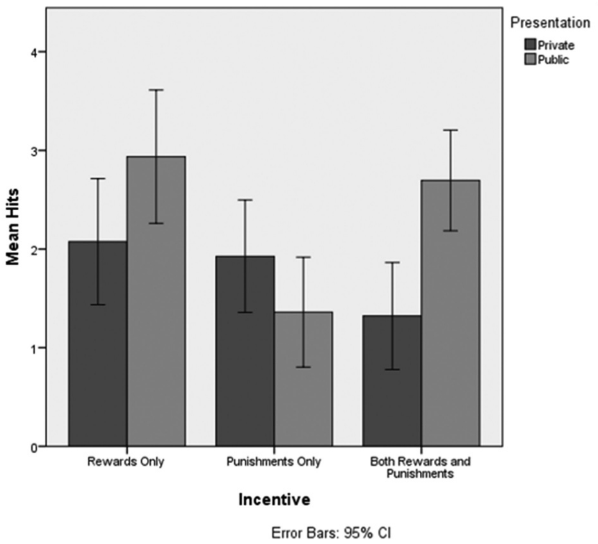


**Figure 4.** Experiment 2: Incentive x presentation interaction on hits.

H4a argued that public presentation would increase motivation to report phishing messages, but no significant effect on motivation was observed. However, there were significant effects from presentation on the number of hits and false positives. Consistent with our hypothesizing, public presentation increased the number of hits and interacted with incentives to increase the number of hits even more. Public presentation also increased the number of false positives. These findings support H4b and H4c. Finally, supplemental analysis (see Online Supplemental Appendix 6) did not uncover negative effects from the leaderboard on work productivity and disruption.

## *Experiment 3: Leaderboards, Training, and Warnings*

To test H5, we gathered participants from a sophomore-level information systems class at a university in the Northeastern U.S. Participants were given course credit for participating in both phases of the experiment. Participants who did not wish to participate were provided with alternative credit opportunities. A total of 385 completed the pre-survey, but only 336 attended the lab sessions. Additionally, 37 participants failed attention checks or did not follow instructions during the lab session and were excluded from data analysis. Therefore, a total of 293 participants completed the experiment. Analysis of excluded participants indicated no threat from non-response bias (see Online Supplemental Appendix 6). Of the participants in our final sample, 44 percent were female, reported a mean age of 20.05 (SD = 2.4), and had 2.51 (SD = .7) years of education after high school. Participants regularly experienced phishing attacks. 47.8 percent of participants reported knowing someone who had fallen for a phishing message, and 23.9 percent of participants reported nearly falling for a phishing message themselves.

## Experiment 3: Independent Variables

To test H5, Experiment 3 consisted of a 2 (leaderboard vs. no leaderboard) x 2 (training vs. no training) x 2 (external warning vs. no external warning) x 2 (phishing warning vs. no phishing warning) factorial experiment. The *leaderboard* displayed points awarded to participants for correctly reporting phishing messages and was visible to all participants in a lab session. Similar to Experiment 2 (see Figure 3, public condition), participants were identified by their intern numbers. They received 100 points for a correct report of phishing messages (hit) and a 25-point deduction for an incorrect report (false positive).

*Training* was performed during the pre-survey of Experiment 3 (before the lab session). It contained a professionally produced video describing the perils of phishing, identifying a phishing message, and what to do if participants received one. Following training, participants were quizzed about the training content, and those who incorrectly responded were excluded from the study.[ii] During the lab session, participants in the *external warning* condition saw a label prepended to the body of emails originating from outside of the organization (see Online Supplemental Appendix 3). This label warned that the message was external and that recipients should not "click links or open attachments unless you recognize the sender and know the content is safe." Participants in the *phishing warning* condition viewed a label at the top of 60 percent of phishing messages warning that the message was possible phishing (see Online Supplemental Appendix 3). A 60 percent accuracy rate was selected because prior research suggests it is low enough to make manual reporting meaningful but high enough so that the warnings still assist participants in discriminating between phishing and legitimate messages [1-3]. When applicable, the orientation participants received included descriptions of the warning conditions and disclosed the 60 percent accuracy rate of the phishing warning.

## Experiment 3: Evaluation

A total of 248 (84.6 percent) participants reported at least one suspicious message as phishing, and the mean number of reports per participant was M = 3.69, SD = 2.76. Table 8 shows the means for each of the dependent variables.

To examine the effect of a leaderboard in the presence of other phishing interventions (H5), we performed a MANCOVA with the presence of leaderboard, training, external warnings, and phishing warnings as independent variables. Motivation, hits, false positives served as dependent variables and propensity to trust, perceived internet risk, computer self-efficacy internal, computer self-efficacy external, and phishing experience as covariates. The complete results of the MANCOVA (including all interaction results from covariates) are reported in Online Supplemental Appendix 6. Significant multivariate tests are shown in

**Table 8.** Means for the dependent variables in Experiment 3.

| Condition | | N | Motivation (SD) | Hits (SD) | False Positives (SD) |
|---|---|---|---|---|---|
| Leaderboard | Leaderboard | 143 | .090 (1.29) | 2.71 (1.83) | 1.45 (1.46) |
| | No Leaderboard | 150 | -.020 (1.19) | 2.15 (1.80) | 1.07 (1.40) |
| Training | Training | 153 | -.004 (1.23) | 2.51 (1.83) | 1.34 (1.42) |
| | No Training | 140 | .080 (1.26) | 2.34 (1.84) | 1.19 (1.45) |
| External Warning | External Warning | 144 | -.040 (1.30) | 2.22 (1.80) | 1.39 (1.44) |
| | No External warning | 149 | .110 (1.17) | 2.26 (1.85) | 1.15 (1.43) |
| Phishing Warning | Phishing Warning | 145 | .040 (1.35) | 2.89 (1.95) | 1.12 (1.45) |
| | No Phishing Warning | 148 | .030 (1.13) | 1.97 (1.59) | 1.41 (1.42) |

Table 9 and reveal small effects for the leaderboard and external warnings and a large effect for phishing warnings. In addition, several significant interactions with small effect sizes were observed involving the leaderboard, including the four-way interaction between all manipulated variables. Univariate tests were performed (Table 10) to interpret the significant multivariate effects, and interactions involving leaderboards were plotted and interpreted.

To interpret the significant univariate tests for the leaderboard, we conducted a series of comparisons with a Bonferroni correction. The significant effects for the leaderboard condition are shown in Table 11.

Additionally, there were numerous interactions between the leaderboard and other phishing interventions. The leaderboard x phishing warning interaction produced an effect on motivation that approached significance and is illustrated in Figure 5. The leaderboard x training interaction and leaderboard x training x external warning x phishing warning interaction significantly affected the number of hits. Since the first interaction affecting hits is subsumed by the second, we focus our attention on the four-way interaction and illustrate the effect of the leaderboard amidst conditions of other phishing interventions (see Figure 6).

**Table 9.** Significant multivariate effects of manipulated variables in Experiment 3.

| Effects | Pillai's Trace | F Score | p Value | $\eta_p^2$ |
|---|---|---|---|---|
| Leaderboard | 0.042 | 3.926 | 0.009 | 0.042 |
| Training | 0.009 | 0.776 | 0.508 | 0.009 |
| External Warning | 0.049 | 4.683 | 0.003 | 0.049 |
| Phishing Warning | 0.132 | 13.729 | <0.001 | 0.132 |
| Leaderboard x Training | 0.044 | 4.100 | 0.007 | 0.044 |
| Leaderboard x Phishing Warning | 0.029 | 2.694 | 0.046 | 0.029 |
| Leaderboard x Training x External Warning x Phishing Warning | 0.053 | 5.002 | 0.002 | 0.053 |

**Table 10.** Univariate effects of manipulated variables in Experiment 3.

| Effects | Dependent Variable | F Score | p Value | $\eta_p^2$ |
|---|---|---|---|---|
| Leaderboard | Motivation | 1.120 | 0.291 | 0.004 |
| | Hits | 11.643 | 0.001 | 0.041 |
| | False Positives | 4.143 | 0.043 | 0.015 |
| Training | Motivation | 1.405 | 0.237 | 0.005 |
| | Hits | 0.117 | 0.733 | <0.001 |
| | False Positives | 0.381 | 0.538 | 0.001 |
| External Warning | Motivation | 0.456 | 0.500 | 0.002 |
| | Hits | 4.695 | 0.031 | 0.017 |
| | False Positives | 2.611 | 0.107 | 0.010 |
| Phishing Warning | Motivation | 0.486 | 0.486 | 0.002 |
| | Hits | 22.637 | <0.001 | 0.077 |
| | False Positives | 2.325 | 0.128 | 0.008 |
| Leaderboard x Training | Motivation | 0.084 | 0.772 | <0.001 |
| | Hits | 10.575 | 0.001 | 0.037 |
| | False Positives | 0.700 | 0.404 | 0.003 |
| Leaderboard x Phishing Warning | Motivation | 2.800 | 0.095 | 0.010 |
| | Hits | 2.713 | 0.101 | 0.010 |
| | False Positives | 0.042 | 0.839 | <0.001 |
| Leaderboard x Training x External Warning x Phishing Warning | Motivation | 1.198 | 0.275 | 0.004 |
| | Hits | 7.699 | 0.006 | 0.028 |
| | False Positives | 0.928 | 0.336 | 0.003 |

**Table 11.** Experiment 3 post-hoc comparisons of leaderboard conditions on motivation, hits, and false positives.

| Dependent Variable | Condition | Comparison | Mean Difference (Condition – Comparison) | Standard Error | Significance |
|---|---|---|---|---|---|
| Motivation[a] | Leaderboard | No Leaderboard | 9.732 | 9.195 | .291 |
| Hits | Leaderboard | No Leaderboard | .672 | .197 | .001 |
| False Positives | Leaderboard | No Leaderboard | .350 | .172 | .043 |

*Note*: Reflects transformed values for motivation.

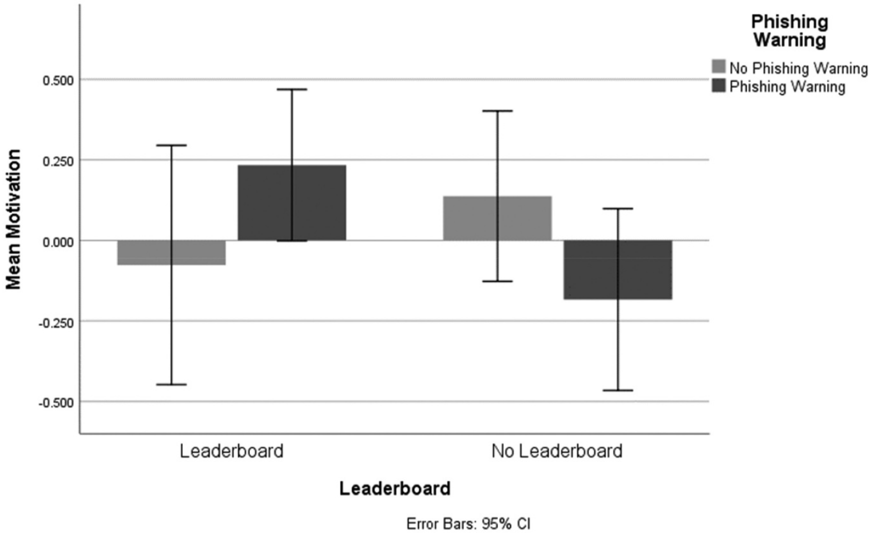


**Figure 5.** Experiment 3: Interaction of leaderboard and phishing warning on motivation.

We did not uncover a main effect from the leaderboard on motivation, which fails to support H5a. Instead, a leaderboard x phishing warning interaction on motivation approached significance, suggesting that the leaderboard's effect on motivation is more nuanced when joined by other phishing mitigation techniques. Motivation was highest when the leaderboard was joined with phishing warnings or when both the leaderboard and warnings were absent. In support of H5b, results revealed a significant main effect from the leaderboard on hits. Furthermore, interactions (particularly with training) were additive with the leaderboard and showed how existing phishing interventions could be layered to increase correct reporting. However, consistent with H5c, the leaderboard also increased the number of incorrectly reported messages. Finally, supplemental analysis (see Online Supplemental Appendix 6) did not uncover negative effects from the leaderboard on work productivity and disruption. However, the leaderboard was implicated in an interaction with external warnings, which lowered work productivity without external warnings but increased work productivity when external warnings were provided.

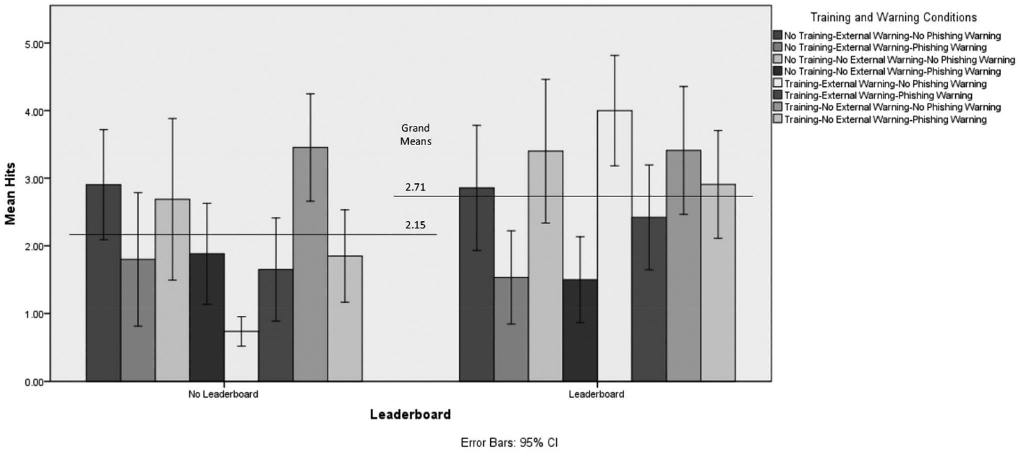


**Figure 6.** Experiment 3: Interaction of leaderboard, training, external warnings, and phishing warnings on the number of hits.

## Discussion

Our research objective was to evaluate the effect of gamification elements and how they can motivate and improve phishing reporting. To achieve this objective, we developed a prototype gamified system using contextualized CET as our kernel theory and tested gamification elements in three laboratory experiments. Our findings support the notion that gamified systems *can* motivate and improve phishing reporting and *can* be combined with other phishing mitigation techniques such as digital training and phishing warnings. Additionally, our findings also demonstrate that the design of the gamification elements is crucial to achieving gains in phishing reporting. Previous gamified systems have produced mixed results [7, 36, 71]. The external rewards that often drive gamification outcomes have led to both positive and negative changes in motivation and task performance [88]. Our results corroborate the delicate balance that exists between gamification elements and outcomes and shed light on the tradeoffs that gamification elements may offer as they influence motivation and performance in phishing reporting. Implications of this research for theory and practice are described in the following section.

### *Implications for Theory*

Among the first implications for theory is that leaderboard conditions across the three experiments produced widespread participation. Across all experiments, 84.5 percent of participants engaged in extra-role behavior and reported at least one suspicious message. Despite the number of work tasks participants were asked to perform, participants also managed to report suspicious messages. As shown by Experiment 3, the presence of a leaderboard increased the number of correctly reported phishing messages while accounting for the effects of other common phishing interventions. This finding demonstrates the unique contribution gamification can make to securing organizations against phishing attacks. Moreover, our results also reveal that gamified systems can be productively combined with other mitigation techniques (e.g., training, warnings) to reduce the phishing threat further. These combinatorial effects corroborate a layered approach to phishing protection whereby no single layer offers complete protection, but multiple layers reduce organizational susceptibility. However, results also suggest that care must be taken in layering mitigation techniques since some combinations result in lower hit rates.

Second, although participants broadly reported suspicious messages, there was variability in the level of influence the gamification elements exerted on experiential and instrumental outcomes. We systematically examined validation, attribution, incentives, and public presentation across two experiments. We discovered that attribution, incentives, and public presentation had the most pronounced sway over motivation, hits, and false positives. Contrary to our expectations, validation's effect on experiential and instrumental outcomes was relatively small. This finding was surprising because validation by an authoritative source is a fundamental component of feedback that serves as a foundation for other gamification elements (e.g., incentives). Because of its essential role, it is likely not possible to disregard validation during any gamification initiative against phishing attacks. Instead, it is more likely that validation need not be publicly included or that it be subsumed in rewards and punishments on leaderboards that are used to motivate and improve reporting performance. Participants appeared much more interested in what they would receive for accurate reporting (incentives) than in whether their judgments were accurate (validation).

The pattern of influential gamification elements is also intriguing because validation (along with private presentation) is most closely associated with intrinsic motivation, while attribution, incentives, and public presentation are most closely associated with external motivation. Since reporting performance results were no different or lower than other treatments associated with attribution, incentives, and public presentation, the treatments associated with intrinsic motivation appeared less prominent. Moreover, the pilot experiment did not reveal damaging effects on intrinsic motivation from attribution, incentives, and public presentation. These findings diverge from the general theory of CET, which acknowledges the power of external rewards but advocates for intrinsic motivation and its effect on performance [26]. Our pattern of findings has two critical implications. First, reporting phishing messages is an activity unlikely to be optimally challenging and self-sustaining through building competence. This implication is consistent with prior research in which phishing detection is considered an ancillary task [61], and students often believe they are already able to detect phishing messages [45]. Second, phishing gamification elements designed to foster intrinsic motivation are unlikely to improve performance. It is more likely that gamification components promoting external motivation improve the ancillary task of reporting phishing messages. However, the robustness of this finding needs to be tested over time. Scholars have observed that engagement with gamified systems changes over time [41, 51], and external motivation may be especially susceptible to these changes.

Next, of the incentives we tested, punishments appeared more potent than rewards. For both motivation and the number of false positives, the introduction of punishments had a negative effect. Even in conditions where rewards and punishments were combined, the level of motivation and false positives were no different from the punishment only condition (and lower than the reward only condition). This finding has important implications for phishing reporting as it sheds light on the relative weight of rewards and punishments. Organizations wishing to incentivize the reporting of phishing messages may provide unbalanced incentives to favor reporting hits (even at the expense of higher false positives).

In Experiments 2 and 3, we adopted this perspective by giving a reward for a hit at four times the punishment of a false positive. Our findings show that punishments carry considerable sway even with such an imbalance as an inhibitor over individuals' motivation and behavior.

Fourth, public attribution and incentives carry unintended consequences. In the binary judgment task of phishing message reporting, incentives and public presentation heightened motivation and altered the judgment thresholds that individuals used to identify phishing. For example, with rewards came more hits, but at the expense of more false positives, and with punishments came fewer false positives, but at the expense of fewer hits. Presentation acted similarly with public display increasing both hits and false positives. These findings corroborate other research in probabilistic deception detection tasks: false positives often increase when individuals are prompted to increase hits [8]. These unintended consequences of gamification elements, together with the imbalance of effects between rewards and punishments, suggest that using gamified systems to externally motivate phishing reporting may undermine performance if the gamification elements are not carefully calibrated. Therefore, care should be taken with gamification elements to ensure appropriate interactions between the gamified system and the user. Such unintended consequences from incentives and public presentation may also partly explain the mixed success of past gamification initiatives.

Previous research [e.g., 51, 73] called for an investigation regarding proper alignment between gamification elements and desired experiential and instrumental outcomes. Our findings suggest that to encourage individuals to contribute to a phishing reporting gamified system, a combination of public attribution and both rewards and punishments are most likely to yield desired gains. This combination of phishing gamification elements is also consistent with neuroscience research suggesting that rewards are more effective in helping individuals take desired actions, and punishments more effectively assist individuals in avoiding undesired actions [72]. This combination meets the demands of increasing motivation and the number of hits. Still, as the results of Experiments 2 and 3 demonstrate, the benefits may come at the expense of elevated false positives. We acknowledge that the amount of rewards and punishments is a critical issue left unanswered by our research. We highlight the investigation of striking this balance as an important area for future research. Our work has shown that simple points assigned by an organization can be meaningful incentives for motivating employees to act. But how many points (or other recognition) an organization should offer for a hit or take away for a false alarm is an open question.

Finally, we detected a minimal drop in work productivity or perceived disruption in robustness checks of the leaderboard manipulations with external warnings combined with leaderboards producing the only detectable negative effect. Additionally, Experiment 2 produced a near significant impact indicating that attribution was functioning to *increase* work productivity, and in Experiment 3, the leaderboard increased productivity when external warnings were absent. These findings suggest that requirements for reporting were not demanding or intrusive. Participants simply needed to forward a suspicious email to report it, a task that took only a few seconds to perform. However, we acknowledge that these findings need to be replicated.

### *Implications for Practice*

With several IT security leaderboard solutions on the market and the absence of research in this area, we offer several important implications for practice. First, we demonstrated proof of concept for security gamification to promote phishing reporting. If organizations wish to broaden and improve reporting of phishing attacks, gamified systems may promise to motivate individuals to participate and incentivize them to report suspicious messages. Furthermore, gamified systems show potential to be used in conjunction with other techniques currently employed in the organizations (e.g., training and warnings).

Next, creating a gamified system that successfully motivates individuals to participate needs not be expensive or complicated, but it needs to foster external motivation to be effective. Attempts to facilitate improvements in phishing reporting via intrinsic motivation are unlikely to be successful because reporting phishing is seen as an extra-role behavior. Simple incentives such as points on a salient website during the experiment session were sufficient to significantly alter experiential and instrumental outcomes in phishing reporting. However, we acknowledge that the validation function necessary to implement incentives could be expensive to put in place if organizations do not have enough IT security resources. For an anti-phishing gamified system to function, an organization would need some way of determining whether reported messages were actually phishing. This would likely require a combination of automated and manual investigative tools and effective reporting, triage, and ticketing systems.

Gamification appears to be a promising approach for creating and strengthening protections against phishing attacks. However, we caution that there can be unintended consequences that appear in response to gamification elements. Careful alignment between gamification elements and intended outcomes is critical. Our research shows that a combination of public attribution with rewards and punishments for contributions is likely to balance demands of widespread participation with accuracy in reporting. However, we acknowledge that organizations will likely have different assessments of the risk that phishing poses. Therefore, we expect that effective incentives will likely vary from one organization to another. Since hits are likely to be more valuable to organizations than the cost of false positives, we anticipate an imbalance between the incentives with more credit given for hits than punishments for false positives. However, the precise amount of incentives will likely be driven by an organizational assessment of phishing risk.

## Limitations and Future Steps

As with all research, this study's limitations are important to keep in mind when considering its implications. First, our study is subject to several limitations common to experimental research. For example, the participants were assigned a role in a fictitious organization. The length of time required for the experiment did not match actual job assignments, which often last much longer than 30 minutes. Although the role would have been familiar to participants, they were also not subject to many of the organizational pressures that actual interns face. These limitations permitted random assignment to conditions and enabled experimental control. But we note them as important areas for future research. Second, we anticipate that the results of this study generalize to the college-aged student population and security contexts. However, additional work is necessary to determine if these results will generalize to a more diverse sample of working adults and other types of tasks. Third, we acknowledge that the function of leaderboards will be altered by the level of attention they receive from organization members and leaders and highlight norms and practices that organizations create around leaderboards as fruitful areas for future work. Also, the consequences some individuals face for being publicly rewarded or punished may be greater than for others, and the privacy issues around public attribution are not addressed. Finally, as previously noted, the individual effects of validation (e.g., correct vs. incorrect feedback) and the amount of rewards and punishments were not examined. Therefore, additional research attention in these areas will likely result in significant contributions to how gamified systems can be deployed to assist organizations in defending themselves against security threats.

## Conclusions

The results of this research support the use of gamification to motivate extra-role security behavior. Individuals in our experiments participated widely in reporting suspected phishing messages. Furthermore, our research demonstrated that gamification elements that are deployed to promote participation dramatically affected individuals' motivation to report suspicious messages and their accuracy in doing so. However, delicate tradeoffs must be balanced in selecting gamification elements. In the context of phishing, the best performing composition of gamification elements included public attribution with rewards and punishments. Gamified systems incorporating these elements provide unique benefits to phishing protection over and above other types of phishing defenses. These findings demonstrate the unique contribution gamification can make to securing organizations against phishing attacks and open the door to a host of new gamification applications in security contexts.

## Notes

[i] The punishment-only condition is analogous to public shaming actions of regulators (e.g., HIPAA data breach wall of shame; https://ocrportal.hhs.gov/ocr/breach/breach_report.jsf) and companies (e.g., Wall of shame for those who fall for phishing attacks; [89]) that try to improve security behavior through threat of punishment. A 2020 survey of UK businesses suggested that 15 percent of respondents name and shame employees for failing mock phishing training (https://www.helpnetsecurity.com/2020/08/05/4-in-10-organizations-punish-staff-for-cybersecurity-errors/). In the context of phishing reporting, there are no rewards for correct reports, just punishment for incorrect reports.

[ii] The training video, developed in part for use in this study, received an honorable mention award from ACM Special Interest Group University and College Computing Sevices which held an international competition for Short Promotional Videos see: https://siguccs.hosting.acm.org/Conference/2016/index.php/awards/

## Acknowledgment

This research was supported by a grant from the National Science Foundation, Social and Economic Division, Project# 1421580.

## Disclosure statement

No potential conflict of interest was reported by the author(s).

## Notes on contributors

*Matthew L. Jensen* (mjensen@ou.edu; corresponding author) is Associate Professor of Management Information Systems and a co-director of the Center for Applied Social Research at the University of Oklahoma. His interests include computer-aided decision making, human-computer interaction, and computer-mediated communication. Dr. Jensen studies how people attribute credibility in mediated interactions and how people filter and evaluate information they find online. His research has been published in *Journal of Management Information Systems, Information Systems Research, MIS Quarterly*, and other journals. He has been the primary investigator or co-primary investigator on externally funded research projects totaling more than $8 million.

*Ryan Wright* (rtwright@virginia.edu) is the C. Coleman McGehee Professor and the Senior Associate Dean of Faculty and Research at the McIntire School of Commerce at the University of Virginia. Dr. Wright's research interests include IT security and privacy, and the diffusion of IT innovations. He has over 70 peer-reviewed publications and has garnered funding from the National Science Foundation, the State of Massachusetts, the State of Virginia. His research has been featured in the *Harvard Business Review, The Washington Post, Forbes, USA Today*, and many other outlets. He has presented his research at such events as TEDx, Salesforce, Personifest, and Association for Finance and Technology.

*Alexandra Durcikova* (alex@ou.edu) is an Associate Professor of MIS and Mertes Presidential Professor at the Price College of Business, University of Oklahoma. She holds a Ph.D. from the University of Pittsburgh. Dr. Durcikova's research focuses on knowledge repositories, knowledge sharing, end-user security, and phishing attack detection. The National Science Foundation funded her research on phishing attack detection. Her publications have appeared in leading journals, including *Information Systems Research, Journal of Management Information Systems, MIS Quarterly, European Journal of Information Systems, Information Systems Journal*, and in the proceedings of numerous international conferences. She has received multiple awards for her teaching.

*Shamya Karumbaiah* (shamya16@gmail.com) is a postdoctoral fellow in the Human-Computer Interaction Institute at Carnegie Mellon University. She earned her Ph.D. from the University of Pennsylvania working as a research fellow at the Penn Center for Learning Analytics. She will be joining the University of Wisconsin-Madison as an assistant professor in Spring 2023. Dr. Karumbaiah's research focuses on promoting student engagement and learning in adaptive learning environments in a fair and equitable manner. Her work has been published in leading journals on affective computing and educational artificial intelligence, with four of her first-authored articles nominated for best research paper awards. For her work on bias, she was selected in the 2021 cohort of EECS rising stars by Massachussetts Institute of Technology (MIT).

## References

1. Abbasi, A.; and Chen, H. A comparison of tools for detecting fake websites. *Computer*, *42*, 10 (2009), 78–86.
2. A. Abbasi, F. Zahedi and Y. Chen, 2012. Impact of anti-phishing tool performance on attack success rates, 2012 *IEEE International Conference on Intelligence and Security Informatics*, Washington D.C., pp. 12–17, doi:10.1109/ISI.2012.6282648
3. Abbasi, A.; Zhang, Z.; Zimbra, D.; Chen, H.; and Nunamaker Jr, J.F. Detecting fake websites: The contribution of statistical learning theory. *MIS Quarterly*, *34*, 3 (2010), 435–461.
4. Allodi, L.; Chotza, T.; Panina, E.; and Zannone, N. The need for new antiphishing measures against spear-phishing attacks. *IEEE Security & Privacy*, *18*, 2 (2019), 23–34.

5. Alsharnouby, M.; Alaca, F.; and Chiasson, S. Why phishing still works: User strategies for combating phishing attacks. *International Journal of Human-Computer Studies*, *82* (2015), 69–82.
6. Amo, L.; Liao, R.; Kishore, R.; and Rao, H.R. Effects of structural and trait competitiveness stimulated by points and leaderboards on user engagement and performance growth: A natural experiment with gamification in an informal learning environment. *European Journal of Information Systems*, *29*, 6 (2020), 1–27.
7. Baxter, R.J.; Holderness Jr, D.K.; and Wood, D.A. Applying basic gamification techniques to IT compliance training: Evidence from the lab and field. *Journal of information systems*, *30*, 3 (2016), 119–133.
8. Biros, D.; George, J.; and Zmud, R. Inducing sensitivity to deception in order to improve deception detection and task accuracy. *MIS Quarterly*, *26*, 2 (2002), 119–144.
9. Bogost, I. Gamification is bullshit. *The Atlantic*, (August 9, 2011). http://www.theatlantic.com/technology/archive/2011/08/gamification-is-bullshit/243338, (Accessed on July 22, 2022).
10. Bogost, I. Persuasive games: Exploitaionware. http://www.gamasutra.com/view/feature/6366/persuasive_games_exploitationware.php., 2011. Accessed January 6, 2020.
11. Bradley, T. Defending local government agencies from rising threat of ransomware. Forbes.com (February 10, 2020). https://www.forbes.com/sites/tonybradley/2020/02/10/defending-local-government-agencies-from-rising-threat-of-ransomware/?sh=7a71a09462e0 . Accessed July 22, 2022.
12. Brown, T.A. *Confirmatory Factory Analysis for Applied Research*. New York, NY: The Guilford Press, 2006.
13. Burke, B. Gamification primer: Life becomes a game. 2011 Gartner Inc. https://www.gartner.com/guest/purchase/registration?resId=1528016&srcId=1-3478922230 . Accessed July 22, 2022.
14. Canfield, C.I.; Fischhoff, B.; and Davis, A. Quantifying phishing susceptibility for detection and behavior decisions. *Human Factors*, *58*, 8 (2016), 1158–1172.
15. Caputo, D.D.; Pfleeger, S.L.; Freeman, J.D.; and Johnson, M.E. Going spear phishing: Exploring embedded training and awareness. *IEEE Security & Privacy*, *12*, 1 (2014), 28–38.
16. Chambers, P. How to deal with repeat cybersecurity offenders. *People Management*, (January 26, 2019). https://www.peoplemanagement.co.uk/article/1743954/how-to-deal-cyber-security-offenders. Accessed July 22, 2022.
17. Chen, Y.; Zahedi, F.M.; Abbasi, A.; and Dobolyi, D. Trust calibration of automated security IT artifacts: A multi-domain study of phishing-website detection tools. *Information & Management*, *58*, 1 (2021), 1–16.
18. Columbus, L. 5 ways machine learning can thwart phishing attacks. Forbes.com, (August 12, 2020) https://www.forbes.com/sites/louiscolumbus/2020/08/12/5-ways-machine-learning-can-thwart-phishing-attacks/?sh=526d3be61035 .Accessed July 24, 2022.
19. Compeau, D.R.; Marcolin, B.; Kelley, H.; and Higgins, C.A. Research commentary—Generalizability of information systems research using student subjects—A reflection on our practices and recommendations for future research. *Information Systems Research*, *23*, 4 (2012), 1093–1109.
20. Constant, D.; Sproull, L.S.; and Kiesler, S. What's mine is ours, or is it? A study of attitudes about information sharing. *Information Systems Research*, *5*, 4 (1994), 400–421.
21. Cristofini, O.; and Roulet, T.J. Playing with trash: How gamification contributed to the bottom-up institutionalization of zero waste. *Academy of Management Proceedings*: Academy of Management 2020, pp. 1–6.
22. D'Arcy, J.; and Teh, P.-L. Predicting employee information security policy compliance on a daily basis: the interplay of security-related stress, emotions, and neutralization. *Information & Management*, *56*, 7 (2019), 1–14.
23. Deci, E.L. Effects of externally mediated rewards on intrinsic motivation. *Journal of Personality and Social Psychology*, *18*, 1 (1971), 105–115.
24. Deci, E.L. *Intrinsic Motivation*. New York, NY: Plenum Press, 1975.

25. Deci, E.L.; and Ryan, R.M. The empirical exploration of intrinsic motivational processes. In L. Berkowitz (ed.), *Advances in Experimental Social Psychology*, New York, NY: Academic Press, 1980, pp. 39–80.
26. Deci, E.L.; and Ryan, R.M. *Intrinsic Motivation and Self-Determination in Human Behavior*. New York, NY: Plenum Press, 1985.
27. Dennis, A.R.; and Valacich, J.S. Conducting experimental research in information systems. *Communications of the Association for Information Systems*, *7*, 1 (2001), 1–41.
28. Deterding, S.; Dixon, D.; Khaled, R.; and Nacke, L. From game design elements to gamefulness: Defining gamification. *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments*, Tampere, Finland: ACM, 2011, pp. 9–15.
29. Dincelli, E.; and Chengalur-Smith, I. Choose your own training adventure: Designing a gamified SETA artefact for improving information security and privacy through interactive storytelling. *European Journal of Information Systems*, *29*, 6 (2020), 669–687.
30. Dur, R.; and Tichem, J. Altruism and relational incentives in the workplace. *Journal of Economics & Management Strategy*, *24*, 3 (2015), 485–500.
31. FBI. Internet Crime Report (IC3). Federal Bureau of Investigation, 2021.
32. Festinger, L. A theory of social comparison processes. *Human Relations*, *7*, 2 (1954), 117–140.
33. Fitz-Walter, Z.; Tjondronegoro, D.; and Wyeth, P. Orientation passport: using gamification to engage university students. *Proceedings of the 23rd Australian Computer-Human Interaction Conference*, Canberra, Australia: ACM, 2011, pp. 122–125.
34. Gartner Research. Gartner Says by 2014, 80 Percent of Current Gamified Applications Will Fail to Meet Business Objectives Primarily Due to Poor Design. Gartner.com, (November 27, 2012), http://www.gartner.com/technology/research/gamification/. Accessed January 6, 2020.
35. Hamari, J. Transforming homo economicus into homo ludens: A field experiment on gamification in a utilitarian peer-to-peer trading service. *Electronic Commerce Research and Applications*, *12*, 4 (2013), 236–245.
36. Hamari, J.; Koivisto, J.; and Sarsa, H. Does gamification work? - A literature review of empirical studies on gamification. *2014 47th Hawaii International Conference on System Sciences (HICSS)*: IEEE, 2014, pp. 3025–3034.
37. Hanus, M.D.; and Fox, J. Assessing the effects of gamification in the classroom: A longitudinal study on intrinsic motivation, social comparison, satisfaction, effort, and academic performance. *Computers & Education*, *80* (2015), 152–161.
38. Hong, W.; Chan, F.K.Y.; Thong, J.Y.L.; Chasalow, L.C.; and Dhillon, G. A framework and guidelines for context-specific theorizing in information systems research. *Information Systems Research*, *25*, 1 (2014), 111–136.
39. Hsu, J.S.-C.; Shih, S.-P.; Hung, Y.W.; and Lowry, P.B. The role of extra-role behaviors and social controls in information security policy effectiveness. *Information Systems Research*, *26*, 2 (2015), 282–300.
40. Hung, S.-Y.; Durcikova, A.; Lai, H.-M.; and Lin, W.-M. The influence of intrinsic and extrinsic motivation on individuals' knowledge sharing behavior. *International Journal of Human-Computer Studies*, *69*, 6 (2011), 415–427.
41. Huotari, K.; and Hamari, J. A definition for gamification: Anchoring gamification in the service marketing literature. *Electronic Markets*, *27*, 1 (2017), 21–31.
42. IBM Security. Cost of a Data Breach Report 2021. https://www.ibm.com/security/data-breach: IBM Security, 2021.
43. Jampen, D.; Gür, G.; Sutter, T.; and Tellenbach, B. Don't click: towards an effective anti-phishing training. A comparative literature review. *Human-centric Computing and Information Sciences*, *10*, 1 (2020), 1–41.
44. Jarvenpaa, S.L.; Tractinsky, N.; and Saarinen, L. Consumer trust in an internet store: A cross-cultural validation. *Journal of Computer-Mediated Communication*, *5*, 2 (1999).
45. Jensen, M.L.; Dinger, M.; Wright, R.T.; and Thatcher, J.B. Training to mitigate phishing attacks using mindfulness techniques. *Journal of Management Information Systems*, *34*, 2 (2017), 597–626.

46. Jensen, M.L.; Durcikova, A.; and Wright, R.T. Using susceptibility claims to motivate behaviour change in IT security. *European Journal of Information Systems*, *30*, 1 (2020), 27–45.
47. Jensen, M.L.; Lowry, P.B.; and Jenkins, J.L. Effects of automated and participative decision support in computer-aided credibility assessment. *Journal of Management Information Systems*, *28*, 1 (2011), 203–236.
48. Kumaraguru, P.; Sheng, S.; Acquisti, A.; and Cranor, L.F. Lessons from a real world evaluation of anti-phishing training. *2008 eCrime Researchers Summit*: IEEE, 2008, pp. 1–12.
49. Kuvaas, B.; Buch, R.; Gagné, M.; Dysvik, A.; and Forest, J. Do you get what you pay for? Sales incentives and implications for motivation and changes in turnover intention and work effort. *Motivation and Emotion*, *40*, 5 (2016), 667–680.
50. Lee, A.S.; and Baskerville, R.L. Generalizing generalizability in information systems research. *Information Systems Research*, *14*, 3 (2003), 221–243.
51. Liu, D.; Santhanam, R.; and Webster, J. Towards meaningful engagement: A framework for design and research of gamified information systems. *MIS Quarterly*, *41*, 4 (2017), 1011–1034.
52. Lowry, P.B.; Petter, S.; and Leimeister, J.M. Desperately seeking the artefacts and the foundations of native theory in gamification research: Why information systems researchers can play a legitimate role in this discourse and how they can better contribute. *European Journal of Information Systems*, *29*, 6 (2020), 609–620.
53. Magellan. Magellan was recently the victim of a criminal ransomware attack. Attorney General, State of California, 2020.
54. Malhotra, N.K.; Kim, S.S.; and Agarwal, J. Internet users' information privacy concerns (IUIPC): The construct, the scale, and a causal model. *Information Systems Research*, *15*, 4 (2004), 336–355.
55. Maltseva, K.; Fieseler, C.; and Trittin-Ulbrich, H. The challenges of gamifying CSR communication. *Corporate Communications: An International Journal*, *24*, 1 (2019), 44–62.
56. McGonigal, J. *Reality Is Broken: Why Games Make Us Better and How They Can Change the World*. New York, NY: Penguin Press HC, 2011.
57. Mekler, E.D.; Brühlmann, F.; Opwis, K.; and Tuch, A.N. Disassembling gamification: The effects of points and meaning on user motivation and performance. *CHI'13 Extended Abstracts on Human Factors in Computing Systems*, Paris, France: ACM, 2013, pp. 1137–1142.
58. Moody, G.D.; Siponen, M.; and Pahnila, S. Toward a unified model of information security policy compliance. *MIS Quarterly*, *42*, 1 (2018), 285–312.
59. Moorman, C.; and Matulich, E. A model of consumers' preventive health behaviors: The role of health motivation and health ability. *Journal of Consumer Research*, *20*, 2 (1993), 208–228.
60. NCC Group. Phishing Mitigations: Configuring Microsoft Exchange to Clearly Identify External Emails. 2016.
61. Neupane, A.; Rahman, M.L.; Saxena, N.; and Hirshfield, L. A Multi-Modal Neuro-Physiological Study of Phishing Detection and Malware Warnings. *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*, Denver, Colorado, USA: ACM, 2015, pp. 479–491.
62. NIST. Awareness, Training, & Education. National Institute of Standards and Technology - Information Technology Laboratory, 2016.
63. Nunamaker, J.F.; Chen, M.; and Purdin, T.D.M. Systems development in information systems research. *Journal of Management Information Systems*, *7*, 3 (1990), 89–106.
64. O'Flaherty, K. How gamification can boost cyber security. Information Age, (February 29, 2019) https://www.information-age.com/gamification-can-boost-cyber-security-123479658/. Accessed July 22, 2022.
65. Oppong-Tawiah, D.; Webster, J.; Staples, S.; Cameron, A.-F.; de Guinea, A.O.; and Hung, T.Y. Developing a gamified mobile application to encourage sustainable energy use in the office. *Journal of Business Research*, *106* (2020), 388–405.

66. Pavlou, P.A.; and Gefen, D. Building effective online marketplaces with institution-based trust. *Information Systems Research*, *15*, 1 (2004), 37–59.
67. Penenberg, A.L. *Play at Work: How Games Inspire Breakthrough Thinking*. New York, NY: Penguin, 2013.
68. Robson, K.; Plangger, K.; Kiesler, S.; McCarthy, I.; and Pitt, L. Game on: Engaging Customers and Employees Through Gamification. *Business Horizons*, *59*, 1 (2015), 29–36.
69. Ryan, R.M. Control and information in the intrapersonal sphere: An extension of cognitive evaluation theory. *Journal of Personality and Social Psychology*, *43*, 3 (1982), 450–461.
70. Ryan, R.M.; Rigby, C.S.; and Przybylski, A. The motivational pull of video games: A self-determination theory approach. *Motivation and Emotion*, *30*, 4 (2006), 344–360.
71. Seaborn, K.; and Fels, D.I. Gamification in theory and action: A survey. *International Journal of Human-Computer Studies*, *74* (2015), 14–31.
72. Sharot, T. What motivates employees more: Rewards or punishments? 2017 Harvard Business Review, (September 26, 2019), https://hbr.org/2017/09/what-motivates-employees-more-rewards-or-punishments. Accessed July 24, 2022.
73. Silic, M.; and Lowry, P.B. Using design-science based gamification to improve organizational security training and compliance. *Journal of Management Information Systems*, *37*, 1 (2020), 129–161.
74. Sophos. Phishing tricks - The top ten treacheries of 2020. Naked Security, 2020.
75. Svrluga, S. Education department warns that students on financial aid are being targeted in phishing attacks. *The Washington Post*: https://www.washingtonpost.com/education/2018/09/15/education-department-warns-that-students-financial-aid-are-being-targeted-phishing-attacks, 2018.
76. Thatcher, J.B.; Zimmer, C.; Gundlach, M.J.; and McKnight, D.H. Internal and external dimensions of computer self-efficacy: An empirical examination. *IEEE Transactions on Engineering Management*, *55*, 4 (2008), 628–644.
77. Thiebes, S.; Lins, S.; and Basten, D. Gamifying information systems-a synthesis of gamification mechanics and dynamics. (2014) *Twenty Second European Conference on Information Systems*, Tel Aviv, Israel, pg. 1-17.
78. Trittin, H.; Fieseler, C.; and Maltseva, K. The serious and the mundane: Reflections on gamified CSR communication. *Journal of Management Inquiry*, *28*, 2 (2019), 141–144.
79. Vallerand, R.J.; and Reid, G. On the causal effects of perceived competence on intrinsic motivation: A test of cognitive evaluation theory. *Journal of Sport and Exercise Psychology*, *6*, 1 (1984), 94–102.
80. Vance, A.; Lowry, P.B.; and Eggett, D. Using accountability to reduce access policy violations in information systems. *Journal of management information systems*, *29*, 4 (2013), 263–290.
81. Vance, A.; Lowry, P.B.; and Eggett, D.L. Increasing accountability through the user interface design artifacts: A new approach to addressing the problem of access-policy violations. *MIS Quarterly*, *39*, 2 (2015), 345–366.
82. Wasko, M.; and Faraj, S. Why should I share? Examining social capital and knowledge contribution in electronic networks of practice. *MIS Quarterly*, *29*, 1 (2005), 35–57.
83. Werbach, K.; and Hunter, D. *For the win: How Game Thinking Can Revolutionize Your Business*. Wharton School Press; Second edition (November 10, 2020).
84. Williams, E.J.; Hinds, J.; and Joinson, A.N. Exploring susceptibility to phishing in the workplace. *International Journal of Human-Computer Studies*, *120* (2018), 1–13.
85. Williams, J.; and Adelson, J. (2020) Ransomware attacks on New Orleans, other Louisiana entities, part of growing trend. The Time-Picayune, (February 9, 2020). https://www.nola.com/news/politics/article_7d22e948-3e31-11ea-98bc-9b69342bc6a8.html (Accessed on July 24, 2022).
86. Wright, R.T.; Chakraborty, S.; Basoglu, A.; and Marett, K. Where did they go right? Understanding the deception in phishing communications. *Group Decision and Negotiation*, *19*, 4 (2010), 391–416.

87. Wright, R.T.; and Marett, K. The influence of experiential and dispositional factors in phishing: An empirical investigation of the deceived. *Journal of Management Information Systems*, *27*, 1 (2010), 273–303.

88. Wu, Y.; Kankanhalli, A.; and Huang, K. Gamification in Fitness Apps: How Do Leaderboards Influence Exercise? (2015). *Proceedings of the 36th International Conference on Information Systems*, Fort Worth, TX, USA. pg. 1-12.

89. Yukl, G.; Wexley, K.N.; and Seymore, J.D. Effectiveness of pay incentives under variable ratio and continuous reinforcement schedules. *Journal of Applied Psychology*, *56*, 1 (1972), 19–23.