frontiers in Education

ORIGINAL RESEARCH
published: 05 January 2022
doi: 10.3389/feduc.2021.807277

Phish Derby: Shoring the Human Shield Through Gamified Phishing Attacks

Matthew Canham[1*†], *Clay Posey*[2†] *and Michael Constantino*[3]

[1]*Beyond Layer Seven, LLC, Oviedo, FL, United States,* [2]*Information Systems, Marriott School of Business, Brigham Young University, Provo, UT, United States,* [3]*Information Security Office, University of Central Florida, Orlando, FL, United States*

OPEN ACCESS

Edited by:
Linas Bukauskas,
Vilnius University, Lithuania

Reviewed by:
Aušrius Juozapavičius,
General Jonas Žemaitis Military Academy of Lithuan, Lithuania
Kristina Lapin,
Vilnius University, Lithuania

****Correspondence:***
Matthew Canham
mcanham@belay7.com

†These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Higher Education, a section of the journal Frontiers in Education

Received: *02 November 2021*
Accepted: *07 December 2021*
Published: *05 January 2022*

Citation:
Canham M, Posey C and Constantino M (2022) Phish Derby: Shoring the Human Shield Through Gamified Phishing Attacks. Front. Educ. 6:807277. doi: 10.3389/feduc.2021.807277

To better understand employees' reporting behaviors in relation to phishing emails, we gamified the phishing security awareness training process by creating and conducting a month-long "Phish Derby" competition at a large university in the U.S. The university's Information Security Office challenged employees to prove they could detect phishing emails as part of the simulated phishing program currently in place. Employees volunteered to compete for prizes during this special event and were instructed to report suspicious emails as potential phishing attacks. Prior to the beginning of the competition, we collected demographics and data related to the concepts central to two theoretical foundations: the Big Five personality traits and goal orientation theory. We found several notable relationships between demographic variables and Phish Derby performance, which was operationalized from the number of phishing attacks reported and employee report speed. Several key findings emerged, including past performance on simulated phishing campaigns positively predicted Phish Derby performance; older participants performed better than their younger colleagues, but more educated participants performed poorer; and individuals who used a mix of PCs and Macs at work performed worse than those using a single platform. We also found that two of the Big Five personality dimensions, extraversion and agreeableness, were both associated with poorer performance in phishing detection and reporting. Likewise, individuals who were driven to perform well in the Phish Derby because they desired to learn from the experience (i.e., learning goal orientation) performed at a lower level than those driven by other goals. Interestingly, self-reported levels of computer skill and the perceived ability to detect phishing messages failed to exhibit a significant relationship with Phish Derby performance. We discuss these findings and describe how focusing on motivating the good in employee cyber behaviors is a necessary yet too often overlooked component in organizations whose training cyber cultures are rooted in employee click rates alone.

Keywords: phishing, cybersecurity awareness training, gamification, NIST phish scale, protective stewards, repeat clickers

INTRODUCTION

Despite significant and increasing organizational spending on cybersecurity technologies and associated efforts, successful threats abound. For example, while organizational leaders are expected to spend more than $150 billion US on cyber and related technologies and services in 2021 (Gartner, 2021), threats related to remote work, cloud adoption, healthcare, and other domains continue to flourish (CheckPoint, 2021). Thus, cyber "solutions" are not always what they appear, and throwing technology at the cyber problem will create rather than solve problems (Schneier, 2015).

An important realization has been that organizational cybersecurity efforts depend largely on the employees who reside within organizational walls. These individuals are central to the effectiveness of organizational actions to protect sensitive assets, and research has shown that they can be detrimental (e.g., sabotage and computer abuse) (Straub and Nance, 1990; Willison and Warkentin, 2013) as well as beneficial (e.g., protective motivated behaviors, precaution taking) (Boss et al., 2009; Posey et al., 2013; Burns et al., 2019) to their employers. Employee actions thus range from accidental errors to malicious acts of sabotage on the negative side and forced compliance to security championing on the positive side.

A specific, significant context where employees continue to affect their organizations is how these individuals respond to phishing attempts that come through corporate email systems. Online phishing is a common attack vector used by external actors to penetrate organizational networks, steal employee credentials, and commit other forms of harm. In fact, more than 90% of malicious software is delivered by email, with personalized phishing attacks (i.e., spear phishing) being the entry gate (Purplesec, 2021). Because of this massive potential for injury, organizations have focused on how best to reduce the risk stemming from employees who encounter and fall victim to phishing attacks. These efforts rely largely on simulated phishing campaigns wherein employees encounter emails that mimic real phishing attacks, and the resulting failure metrics are used to examine progress within an employee base.

Notwithstanding the importance of assessing the number of employees who fall victim to these mock attacks, it is important to note how employees can also have positive reactions to phishing attacks—reactions that alert organizational representatives to the potential threat (Canham et al., 2021). It is unfortunate that many of these positive reactions are often overshadowed by the failures (i.e., successful mock attacks) despite serving as an important warning signal or beacon to the organization that something could be wrong. At a time when cybersecurity remains a top priority for leadership, but funding for the requisite resources is unable to keep pace with the ever-evolving threat landscape, it would serve organizations' interests to also provide significant focus on the positive spectrum of employees' cyber behavior.

To increase our understanding of this phenomenon, which we refer to as the "protective steward phenomenon," we gamified a series of simulated phishing campaigns to see how such an alteration would influence employee cyber behaviors. Gamification refers to the "use of computer games and features of games for non-game purposes" (Fleming et al., 2020, p. 2). These campaigns, collectively called a "Phish Derby" competition, allowed employees to compete against one another in their efforts to detect and create an alert when encountering simulated phishing emails.

Given the evidence showing how the gamification of learning-based exercises can increase participants' engagement and overall learning (Marín et al., 2018; Groening and Binnewies, 2019), we explored whether and how gamification could be used to foster positive employee reactions and experiences with a form of training (i.e., simulated phishing campaigns) seen by some workers as a source for decreased productivity and increased levels of boredom, anxiety, stress, embarrassment, and even ostracism (Conley, 2021; Emm, 2021; Ferrell, 2021). At the very least, gamification could prove to increase user attentiveness during these activities, which could then possibly translate to better performance during real attacks. In addition, not only was correct identification of phishing attempts important, but given the need for organizations to be able to respond to threats as quickly as possible, employee response times (i.e., time difference between phish receipt and employee alert) were also tabulated. Therefore, our experiment with the Phish Derby and its associated results provides a more holistic view to positive employee behaviors regarding one of the most harmful attack vectors used against modern organizations—online phishing attacks.

BACKGROUND ON PHISHING

Since online phishing and its variants, like business email compromise, continue to be successful attack vectors, especially during the COVID-19 pandemic when cyberattacks increased by 600% (Purplesec, 2021), it is no wonder that substantial scholarly attention has been given to phishing attack detection. Unfortunately, when compared to automated, technical-detection solutions, research on human-based detection efforts is more limited and focuses on how training techniques can be leveraged to enhance detection capabilities (Khonji et al., 2013; Zielinska et al., 2014; Wash and Cooper, 2018). Fortunately, research shows some promise in increasing human-detection capabilities via phishing training embedded directly into corporate email systems (Kumaraguru et al., 2007), but even then, employees might not even fully read or pay attention to the training (Caputo et al., 2013).

Complementing the research on human-detection capabilities, recent efforts have drawn attention to all potential employee behavioral responses to email phishing attacks (Canham et al., 2021). By analyzing the responses of more than 6,000 employees at a large U.S. university over the course of 20 phishing training campaigns and 19 months, this effort demonstrated that a small subset of users (6% of the total population of users) were responsible for repeated phishing training failures (i.e., "Repeat Clickers") and a larger subset (33%) of users ("Protective Stewards") were responsible for reporting these emails to the Information Security Office. Thus, more employees alert their organizations about potential attacks than succumb to phishing

attacks. Unfortunately, this positive-oriented and more sizable employee subpopulation has received relatively limited attention when compared to its smaller and more detrimental counterpart—a concerning trend when so many information security offices are struggling to handle day-to-day operations with limited resources.

One potential way to continue to increase employees' 1) ability to detect and 2) motivation to report phishing emails might be through the gamification of the mock phishing campaign experience. The addition of gaming elements to non-gaming situations in this and other cyber-related contexts has been explored (Francia et al., 2014; Gjertsen et al., 2017; Emm, 2021; Khando et al., 2021). For example, gamification has demonstrated promise in the education of normal users regarding password security (Scholefield and Shepherd, 2019), and gamified systems can increase motivation to comply with security policy and reduce mock phishing failures, significantly outperforming training provided via email (Silic and Lowry, 2020). Different variations of gamification capabilities have also been examined in the context of employees' online self-disclosure (Dincelli and Chengalur-Smith, 2020) and corruption behaviors (Baxter et al., 2017). In addition, previous work on gamified systems have relied on both monetary and non-monetary rewards to incentivize participants (Lewis et al., 2016; Karac and Stabauer, 2017; Meixner et al., 2020; Ueyama et al., 2014). It is evident that gamification can be a useful tool in educating and motivating individuals in a variety of contexts.

Given this opportunity, we extended previous research efforts by exploring the factors surrounding employees who actively choose to alert their information security office when they suspect a rogue email in their inbox. In addition, we wanted to determine if employee response times could be incentivized through such gamification. Akin to the field of positive psychology (Seligman and Csikszentmihalyi, 2014), our goal here is to help motivate positive behaviors rather than correct negative actions and understand whether gamification is a fruitful avenue for this objective.

Possible Employee-Performance Factors

To better understand potential variance in employee performance during our Phish Derby, we relied on concepts found in two theoretical foundations. The first foundation is commonly referred to as the "Big Five" personality traits. These traits include extraversion, emotional stability, agreeableness, conscientiousness, and openness to experience (Norman, 1963; McCrae and Costa, 1987). Because so much has been written on these traits, we briefly discuss them here.

Openness is a trait aligned with intellectual curiosity, creativity, and a preference for novelty. Individuals high in conscientiousness tend to be organized, self-disciplined, and have a need for achievement, whereas individuals high in extraversion tend to be socially outgoing, energetic, and seek stimulation. Agreeableness refers to those who tend to be cooperative, helpful, and well-tempered. Finally, individuals exhibiting neuroticism tend to be prone to anxiety and stress, easily experience unpleasant emotions, and be insecure.

The Big 5 has been examined as an influential factor in studies of information security previously (Pattinson et al., 2012; Uebelacker and Quiel, 2014; Halevi et al., 2015; Welk et al., 2015; Lawson et al., 2017; Sudzina and Pavlicek, 2017); however, how these traits influence phishing susceptibility is not always obvious. For example, people high in conscientiousness might be less susceptible to phishing attempts (Lawson et al., 2017), but they might also be leveraged to help an attack become more likely to succeed (Halevi et al., 2015). Regarding phishing vulnerability, research on individuals high in extraversion has shown mixed results. Two studies have shown increased susceptibility to phishing (Welk et al., 2015; Lawson et al., 2017), while another study (Pattinson et al., 2012) showed a better ability to detect phishing emails. Despite these differences, we believe that one or more of the Big 5 components could play an important role in understanding potential differences in our participants' performance, especially given our unique context of gamification and the inclusion of relatively difficult-to-detect phishing emails in our Phish Derby.

Goal orientation theory (GOT) serves as our second theoretical foundation. This theory explains the reasons why individuals are driven to certain outcomes in achievement-focused tasks. Generally, individuals approach and engage in achievement tasks because they desire to 1) learn (i.e., learning), 2) prove their performance abilities (i.e., prove performance), and/or 3) avoid negative judgments and perceptions of inferiority (i.e., avoid performance) (Brett et al., 1999; Kaplan and Maehr, 2007).

GOT has been used in examining individuals in numerous achievement-focused scenarios. For example, goal orientation concepts have been linked to academic performance, even mediating the relationship between intrinsic motivation and performance (Cerasoli and Ford, 2014). Learning orientations have been linked to expatriates' academic and social adjustment outcomes (Gong and Fan, 2006), and both learning- and performance-orientation goals have shown relationships with team adaptability when facing adversity (Porter et al., 2010). Finally, research has shown that trait-forms of goal orientation explain employee job performance above and beyond cognitive ability and even the personality variables mentioned above (Payne et al., 2007). Determining whether and how these goals drive Phish Derby performance in general, and in comparison, with the "Big Five" personality traits should prove fruitful.

GAMIFIED APPROACH

Gamification was achieved *via* our "Phish Derby" by having participants prove their ability to spot phishing attacks and earn points based upon the number of attacks they successfully reported, as well as how quickly those alerts were issued. To help increase the amount of variance in user responses, the research team utilized very difficult simulated phishing attacks. The KnowBe4 platform was used for the Phish Derby. Participants received monetary prizes (i.e., Amazon gift cards) at the end of the competition, and they also knew that the research team would debrief all who were interested in an online seminar.

TABLE 1 | Email template phish scale difficulty with click and report rates.

Template	Campaign	Number of cues	Premise alignment	Difficulty rating	Click-rate (%)	Report-rate (%)
LinkedIn–People Are Looking at your Profile	1	4 (Few)	Low	Moderately Difficult	8	10
UPS Label Delivery	2	9 (Some)	Medium	Moderately Difficult	18	69
Test of the Notification System	3	11 (Some)	High	Very Difficult	6	67
Sarah Butler Sent You a Secure File	4	8 (Few)	High	Very Difficult	0	73
Knightro's Halloween Costume	5	7 (Few)	Medium	Very Difficult	1	20
COVID-19 Reported Cases in Your Area	6	6 (Few)	High	Very Difficult	3	56

Potential participants were notified of the Phish Derby a week prior to its beginning through email communication. Participation in the Phish Derby was voluntary, and competitors were instructed that because this was a competition, the simulated phishing emails that they received would be more difficult than the regular training emails that they had received in the past. Information Security Office staff informed volunteers that performance during this Phish Derby would not negatively impact their training requirements (e.g., being required to complete additional training if they fell for a simulated phishing message sent as part of this Phish Derby).

A total of six simulated phishing email templates were utilized for the Phish Derby competition. These six were titled "LinkedIn–People Are Looking at your Profile," "UPS Label Delivery," "Test of the Notification System," "Sarah Butler Sent You a Secure File," "Knightro's Halloween Costume," and "COVID-19 Reported Cases in Your Area." The "LinkedIn–People Are Looking at your Profile" template purported to notify the recipient that their profile had been viewed and included a hyperlink that falsely claimed to redirect to LinkedIn. "The UPS Label Delivery" template used the pretext of a UPS delivery notification with a hyperlink made that appeared to redirect to UPS. The "Test of the Notification System" template claimed to be a notification test and requested the receiver verify their contact information through a deceptive hyperlink. The "Sarah Butler Sent You a Secure File" template appeared to be a shared document from Sarah Butler, a fictitious university employee. The "Knightro's Halloween Costume" used the pretext of an invitation to enroll in a university costume contest. The final template, "COVID-19 Reported Cases in Your Area," used the pretext of discovering reported COVID-19 cases in the area through a deceptive linked portal.

Developing an objective metric of email difficulty is a challenge that the NIST Phish Scale seeks to address. This difficulty scale considers two factors in operationalizing phishing email difficulty: first, the number of phishing "cues," and second, the email premise alignment with user role (Steves et al., 2020). These factors were derived from previous empirical work demonstrating their central role in phishing email detection (Greene et al., 2018). Cues refer to inconsistencies within, or characteristics of, the message that may alert the target that the message might be a phishing attempt. Examples of cues include spelling and grammatical errors, technical indicators (e.g., a hyperlink mismatch), odd language, and the use of time pressure. Premise alignment refers to the degree to which the message aligns with the recipient's job role and alludes to the user's context in evaluating the message. Prior research demonstrates the more highly the message premise aligns with the target's job role (i.e., a past-due notice sent to the accounting department, or a resume sent to a human-resources department), the less likely people are to notice detection cues in the message (Greene et al., 2018). We applied the NIST Phish Scale of email difficulty to each of the six simulated phishing email templates that we employed in the Phish Derby, and the difficulty ratings for each template are summarized in **Table 1**.

EXPERIMENTAL METHOD

In early October 2020, participants completed an initial survey that covered demographic and model variables used in our analyses. A total of 116 individuals took part in the initial survey, but attention-check items indicated that only 101 individuals should remain in the study. These individuals then received six simulated phishing emails to their work email address throughout the remainder of October. Participants explicitly agreed to not use any means (technical or otherwise) that would prohibit a fair competition. Any evidence suggesting use of such methods would result in immediate disqualification from the competition. Participants were instructed to report emails as potential phishing attacks by using an embedded "Phish Alert" button as provided by KnowBe4 or by forwarding the email as an attachment to the Security Incident Response Team (SIRT). All interaction with phishing emails (e.g., email receipt, reporting) took place in the 8:00 am—5:00 pm (participants' local time) window. The mean age of our sample was 44.4 years, with 40% identifying as female. Twenty-five percent of our sample was in administrative positions, and 10% was in an IT/IS role.

Our research team collected the number of phishing alerts/reports received from participants as well as the timeliness with which those alerts/reports were received. All participants began the competition with 10 "Derby Bucks." For every simulated phishing email not alerted within 4 h of receiving the email, 1 Derby Buck was subtracted from their total. If the competitor reported the email but only after falling victim to the phishing email, $0.75 was subtracted from their total. This option was available to highlight the fact that while succumbing to a phishing attack is a negative event, it is still of benefit to the organization to report it as soon as possible. No Derby Bucks were removed when participants accurately alerted SIRT within 4 h of receiving the phishing email. At the end of the Phish Derby, competitors

received the following rewards based on the total amount of Derby Bucks remaining in their possession:

> Competitors with at least 6.00 Derby Bucks were awarded a $5.00 Amazon gift card.
>
> Competitors with at least 7.50 Derby Bucks were awarded a $7.00 Amazon gift card.
>
> Competitors with at least 8.50 Derby Bucks were awarded a $10.00 Amazon gift card.

The dependent variable in our model was a performance score normalized for report timeliness against the average response time for each phishing campaign. This represents the importance of timeliness in reporting potential threats to SIRT. For example, if two participants correctly identified all six phishing campaigns, their initial performance score would equal 6.00, but because their average response times differed, the final performance scores would be adjusted relative to those response times. Thus, instead of both participants receiving the same 6.00 performance score, the faster responder might receive a 5.93 and the slower one a 5.78. Thus, the faster the response, the higher the score (assuming the same number of phishing campaigns was identified). This normalized score was not used in the assignment of Derby Bucks mentioned above due to university institutional review board (IRB) stipulations. Participants were not informed of their performance relative to other competitors either during or after the Phish Derby; however, they were informed of the overall Phish Derby detection and reporting performance after the competition had concluded.

MEASURES

In addition to the demographic variables, we used previously published and validated scales to capture our constructs related to our two theoretical foundations. The Big 5 Personality dimensions were assessed using the IPIP–NEO–60 scale (Maples-Keller et al., 2019). This scale employs 60 items to infer an individual's placement along each of the five dimensions on the Five-Factor Personality Scale. Learning (5 items), prove performance (4 items), and avoid performance (4 items)—concepts from GOT—were measured using the 13-item goal orientation scale (Brett and VandeWalle, 1999).

RESULTS

We performed a hierarchical regression analysis where we focused on participants' demographic variables first and then assessed components related to the Big Five personality traits and GOT. Given that we operationalized these components with previously validated measures, each exhibited adequate internal consistency metrics ($\alpha \geq 0.70$). While our sample size is relatively small ($n = 101$), our statistical power ($1\text{-}\beta > 0.99$) did not prohibit us from discussing non-significant relationships. What is rather interesting is that such a relatively simple model

TABLE 2 | Interconstruct correlations.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1. Phish Proneness %																	
2. Previous Phish Received/ Reported %	−0.027																
3. Norm. Performance Score	−0.125	0.538**															
4. Derby emails reported	−0.153	0.544**	0.958**														
5. Average response time	−0.052	−0.020	−0.441**	−0.078													
6. Sex	−0.145	−0.029	0.041	0.026	−0.023												
7. Age	0.123	−0.138	−0.063	−0.077	−0.052	−0.115											
8. Manager	−0.024	0.079	−0.043	−0.017	0.229*	0.140	0.031										
9. IS/IT role	−0.143	0.293**	0.241*	0.227*	−0.108	0.263**	−0.222*	0.194+									
10. Extraversion	0.030	−0.080	−0.241*	−0.196*	0.138	0.009	0.027	0.094	−0.115								
11. Agreeableness	0.152	−0.079	−0.146	−0.136	−0.014	−0.017	0.288**	0.050	0.098	0.282**							
12. Conscientiousness	0.098	−0.094	−0.123	−0.068	0.163	−0.197*	0.089	−0.050	−0.049	0.450**	0.354**						
13. Neuroticism	−0.079	0.077	0.113	0.055	−0.157	−0.135	−0.109	0.018	0.025	−0.441**	−0.453**	−0.432**					
14. Openness	−0.057	−0.022	−0.047	0.008	0.143	0.019	−0.046	−0.045	−0.078	0.186+	0.124	−0.052	−0.167+				
15. Performance intent	−0.214*	0.047	0.074	0.094	0.083	0.023	−0.191+	0.077	0.001	0.167+	0.055	0.120	−0.108	0.159			
16. Learning	−0.227*	−0.097	−0.235*	−0.166+	0.260*	0.228*	−0.056	0.125	0.054	0.283**	0.105	0.176+	−0.365**	0.366**	0.231*		
17. Prove performance	−0.007	−0.167	−0.179+	−0.187+	−0.007	0.160+	−0.119	0.012	−0.012	−0.102	−0.070	−0.140	0.213*	0.004	0.156	0.072	
18. Avoid performance	−0.010	−0.092	0.000	−0.050	−0.178	−0.168+	0.024	−0.076	−0.151	−0.231*	−0.284**	−0.365**	0.418**	−0.123	−0.094	−0.469**	0.430**

+$p < 0.10$; *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

TABLE 3 | Results from hierarchical regression of normalized performance on demographics and predictors.

Variables	Step 1: Demographics		Step 2: Predictors	
	Standardized beta	p-value	Standardized beta	p-value
PhishProneness%	–0.165	0.077	–0.180	0.061
prevPhishRecRep%	0.484	0.000	0.395	0.000
Age	0.139	0.155	0.248	0.018
Education	–0.211	0.030	–0.166	0.087
Tenure	–0.117	0.274	–0.200	0.073
Mac Only	0.064	0.506	0.070	0.483
PC-Mac Mix	–0.191	0.033	–0.200	0.022
Personality				
Extraversion			–0.181	0.066
Agreeableness			–0.261	0.016
Conscientiousness			–0.085	0.407
Neuroticism			–0.140	0.213
Openness			0.055	0.558
Goal Orientation				
Learning Goal			–0.309	0.007
PerfProveGoal			–0.045	0.667
PerfAvoidGoal			–0.186	0.131
Performance Intention			0.158	0.108
R^2	37.7%	0.000	52.0%	0.000
R^2_{adj}	32.5%	0.000	41.9%	0.000

produced a rather large amount of variance in participants' performance (i.e., R2 = 52.0%). The mode for phishing emails reported across all participants was 4, and the average response times (in minutes) for the email templates were 97.0, 66.4, 121.3, 189.0, 37.8, and 77.2 for templates 1–6, respectively. The mean normalized performance score (possible range 0–6) was 2.55 during the Phish Derby. **Table 2** displays the correlations among our variables, and **Table 3** displays our statistical results.

In addition to the large R^2 value, we see several notable relationships between demographic variables and Phish Derby performance. First, participants' exposure to, and performance during, previous simulated phishing campaigns matter as demonstrated by the significance of the percentage of reports relative to phishing emails received by the employees before entering the Phish Derby. We also see that age becomes a significant variable in our analysis, and the positive beta indicates that older participants performed better in the Phish Derby than did their younger counterparts. On the other hand, the years of education had the opposite effect on performance; more years of education led to poorer performance. Finally, in our assessment of whether participants used PCs or Macs at work, or a mix of both, we found that individuals who use a mix performed worse than those using a single platform.

Outside the demographic variables, we see variables of significance within our two theoretical foundations. First, within the Big Five personality dimensions, two personality dimensions influenced performance: extraversion and agreeableness. Extraversion was negatively associated with phishing email reporting performance ($\beta = -0.181$, $p = 0.066$). This finding is interesting because previous research on those high on the extraversion dimension has been mixed. At least three studies have found increased susceptibility to phishing in those higher in extraversion (Welk et al., 2015; Lawson et al., 2017; Anawar et al., 2019), while another study (Pattinson et al., 2012) showed a better ability to detect phishing emails. Messages that utilize likability as a social influence principle have also been found to be more persuasive to people high in extraversion (Alkış and Temizel, 2015). This aspect of personality needs to be examined more because more extraverted individuals might make more attractive targets for criminals, by virtue of having more connections and thus having more connections to target if their account is compromised. Agreeableness has been positively associated with self-reported cybersecurity behaviors in previous research (McCormac et al., 2017; Shappie et al., 2020); however, we found that higher agreeableness was also associated with poorer phish reporting performance in the Phish Derby ($\beta = -0.261$, $p = 0.016$).

From a goal orientation perspective, we found that participants whose goal was overall learning performed significantly worse than those who identified their goal as performing well in the competition. In other words, those who are trying to better themselves at identifying phishing attacks performed at a lower level than those who cared little about overall learning as their main goal. Goals of "performance proving" and "performance to avoid disapproval" did not exhibit a significant relationship with overall performance.

Finally, participants' intention to do well in the Phish Derby was close to becoming a significant component in the model but ultimately was not. In the case of the gamified Phish Derby, intentions were not significantly related to performance—perhaps a case of the "knowing-doing gap" (Workman et al., 2008).

Regarding performance on the various templates, most had relatively low click-rates, with the exception of the UPS Label Delivery email template (18% click-rate). This template received the second highest reporting rate (69%) suggesting that it was among the more interactive templates of the Phish Derby. These results are summarized in **Tables 4**, **5**. While the

TABLE 4 | Click-rates by email template and job role.

	LinkedIn notification	UPS label delivery	Test of notification system	Secure file delivery	Halloween costume contest	COVID-19 in your area
Overall Click-Rate	8%	18%	6%	0%	1%	3%
Administrative	5	7	3	0	0	1
Information Technology	1	2	0	0	0	1
Management	0	1	1	0	0	0
Professor–Instructor	2	8	2	0	1	1
Total	8	18	6	0	1	3

TABLE 5 | Report-rates by email template and job role.

	LinkedIn notification	UPS label delivery	Test of notification system	Secure file delivery	Halloween costume contest	COVID-19 in your area
Overall Report-Rate	10%	69%	67%	73%	20%	56%
Administrative	6	34	32	36	8	29
Information Technology	1	7	5	4	1	4
Management	0	7	7	7	1	7
Professor - Instructor	3	22	24	27	10	17
Total	10	70	68	74	20	57

"Secure File Delivery," "COVID-19 in Your Area," and "Test of the Notification System" were also highly reported, the "Secure File Delivery" received zero clicks. This may have been due to similar simulated phishing emails having been previously used in training. No significant interactions (in click-rates or report-rates) between the email template types and job role were observed. Overall, the click-rates were relatively low compared to previously observed campaigns (Canham et al., 2021). This might have been the result of a self-selected sample of participants with knowledge of their contest participation.

Phish Derby Participant Comments

In addition to our quantitative assessment, we wanted to determine whether the participants viewed the Phish Derby experience as a success. Fortunately, the comments from participants during the debrief regarding the Phish Derby were overwhelmingly positive, and a sample of them includes the following direct quotes:

> "I enjoyed taking part in the phishing derby–seriously a great idea!"
>
> "It was kind of scary though... I would usually delete, but during this I felt like maybe I should report more."
>
> "After I got caught on the first one, I was much more alert for the rest"
>
> "This was a great way to heighten awareness and learn about different kinds of things to watch for."
>
> "I thought I was already aware so kinda wanted to test myself."
>
> "More cautious now."
>
> "When can we do it again?"
>
> "This was a perfect strategy: educational and fun!"
>
> "It was a great learning experience. Thank you."
>
> "(I) would love to see this again thank you"
>
> "I appreciate the Derby tests and will stay vigilant!"

DISCUSSION

We aimed to assess whether the gamification of mock phishing exercises would be successful and whether key factors explaining participant performance would emerge. Further, our goal was to highlight employees' positive, phish-reporting behaviors rather than focus on failure (i.e., click) rates. Our results suggest that gamification can be a useful, interesting, and perhaps even exciting approach to employ in mock phishing exercises—exercises that are usually thought to be intrusive or a waste of time by many employees. Moreover, we were able to determine considerable differences in Phish Derby performance, indicating that some employees are or could become star performers or champions for organizations' security teams in the quest to quickly identify phishing attempts once they clear technical filters.

From a theoretical standpoint, it was interesting to discover that both extraversion and agreeableness exhibited negative relationships with Phish Derby performance. In phishing-susceptibility research, findings relative to the extraversion personality trait have been somewhat mixed, but at least three studies have found that extraverts exhibit increased susceptibility to phishing emails (Welk et al., 2015; Lawson et al., 2017; Anawar et al., 2019). Our findings suggest that extraverts perform more poorly on the positive-oriented behaviors of reporting as well. On the other hand, agreeableness has shown positive associations with self-reported cybersecurity behaviors in previous research (McCormac et al., 2017; Shappie et al., 2020), which is at odds with our results. We do not fully understand why such is the case from our Phish Derby exercise, but it is possible that individuals

who desire to be helpful also do not like to have external pressures to do so. Perhaps the additional pressure of the research team tracking and rewarding response times backfired with participants high in agreeableness. This aspect of our findings and rationale deserves future attention.

An additional finding that surprised us was that the first of the three goal orientations (i.e., learning orientation) was negatively related to participants' performance. In fact, learning-oriented individuals exhibited higher average response times, meaning that they were slower to report suspected phishing attacks to the organizational representatives. Like those high in agreeableness discussed above, it is likely that learning-oriented individuals wish to devote a reasonable amount of time to delve into an issue, and quick-reaction contexts do not bode well for these people. Conversely, these individuals are likely those participants who attended and actively participated in the Phish Derby debrief to learn of overall response rates and to discuss the possible cues within each of the mock phishing templates. Unfortunately, we did not assess this possibility.

Another surprising finding was that more education was related to poorer performance in the Phish Derby. On the one hand, it could be that those with more education perceived themselves to be more capable than those with less education at identifying phishing threats, thereby presenting a situation of overconfidence. On the other hand, a potential explanation for this finding may be the more highly educated participants might also have had a higher workload and/or received substantially more emails on a daily basis than the less educated participants. This finding deserves more investigation in future studies.

We also found that employees who alternate between PC and Mac systems in their daily job tasks performed worse than those using a single platform. Our rationale leans toward multi-platform users having an increased cognitive load due to the switching between platforms, icon sets, and other platform-dependent idiosyncrasies. If they are subconsciously expending mental energy on the effective utilization of the various platforms, they may not have the same level of energy available to recognize more difficult phishing threats. Of course, this assertion deserves more attention and could make for a very interesting experiment.

Several practical implications also emerged from our Phish Derby exercise. First, some CISOs might wonder whether exposing employees to simulated phishing campaigns works, and whether they should pay attention to individual metrics. In the case of gamified competitions, the answer is "yes." Competitors' previous reporting behaviors as a percentage of the total number of phishing campaigns they previously received (made available in the KnowBe4 platform) significantly related to their overall performance (i.e., correct identification and timeliness of report) during the Phish Derby.

Second, anecdotal evidence suggests that some CISOs and other security administrators feel unable to get and maintain older employees' interest and attention regarding security matters. We did not find this to be the case during the Phish Derby. In fact, we found that the more aged employees performed better than the younger employees. Perhaps the friendly competition provided by the Phish Derby and the focus on protective reporting—a positive focus—appealed more to older employees than would a focus on thwarting phishing failures—a negative focus. We believe that research in the positive psychology movement (Seligman and Csikszentmihalyi, 2014) could help provide additional rationale for why such is the case on this matter.

Third, the individuals who really desired to learn about how to improve their future performance on reporting phishing emails were the ones who performed worse in the Phish Derby. Those individuals who tend to join competitions to show themselves and others how good they are and those who try to do well in achievement tasks to avoid negative judgment did not perform differently than those who do not. Thus, a true learning goal orientation affects performance during gamified phishing competitions. This is not to say that these individuals should not be involved in gamified phishing exercises; rather, organizational leaders should aim to provide a meaningful and engaging debrief that is focused on the needs of these employees. These will be the individuals most interested in understanding the cues and contexts that made the phishing templates difficult to detect.

Finally, of the variables that failed to exhibit a significant relationship with Phish Derby performance, two of the most interesting ones were self-reported computer skills and perceived ability to detect phishing messages. These variables never approached statistical significance in our analyses; thus, they were not included in our findings table. But the lack of findings here indicates that individuals who believe that they can identify phishing emails better than others failed to perform any differently from those who did not. Such is the same with self-reported computer knowledge and skills.

CONCLUSION

We complemented the online phishing training exercises of employees at a large U.S. university with a month-long gamification experiment. The gamified experience focused on the positive reporting behaviors of participants rather than click rates alone. In addition, we assessed the speed with which participants reported the simulated phishing emails to the appropriate organizational representatives. As evidenced by the findings and the comments provided by Phish Derby participants during the debrief, we view the gamification effort a success. Moreover, despite low enrollment—partially due to the lack of extensive marketing channels between the Information Security office and employees—we believe this event provided the single-most positive experience between the security office and the university's employees. We encourage CISOs who are looking to improve employee participation in phishing exercises to strongly consider adding gamification elements to their efforts, and we implore other researchers to explore gamification's influence more fully in increasing and sustaining individuals' motivation to serve as stewards of security.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Institutional Review Board, University of Central Florida. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

MCa developed experimental materials, conducted data analysis and wrote portions of the introduction, method, results, discussion, and conclusion. CP developed experimental materials, conducted data analysis and wrote portions of the introduction, method, results, discussion, and conclusion. MCo developed experimental materials, ran the study protocol, and wrote portions of the introduction, method, results, discussion, and conclusion.

FUNDING

Funding for this research was made possible through the University of Central Florida's Office of Research. Funding was also provided by the National Institute of Standards and Technology (NIST) under Financial Assistance Award Number: 60NANB20D189. The views and conclusion contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of NIST or the U.S. Government.

ACKNOWLEDGMENTS

The authors would like to thank the Information Security Office at the University of Central Florida for engaging in a collaborative partnership with faculty. Without such efforts, this research would not have been possible.

REFERENCES

Alkış, N., and Temizel, T. T. (2015). The Impact of Individual Differences on Influence Strategies. *Personal. Individual Differences* 87, 147–152.

Anawar, S., Kunasegaran, D. L., Mas'ud, M. Z., and Zakaria, N. A. (2019). Analysis of Phishing Susceptibility in a Workplace: a Big-Five Personality Perspectives. *J. Eng. Sci. Technol.* 14 (5), 2865–2882.

Baxter, R. J., Holderness, D. K., Jr, and Wood, D. A. (2017). The Effects of Gamification on Corporate Compliance Training: A Partial Replication and Field Study of True Office Anti-corruption Training Programs. *J. Forensic Account. Res.* 2 (1), A20–A30. doi:10.2308/jfar-51725

Boss, S. R., Kirsch, L. J., Angermeier, I., Shingler, R. A., and Boss, R. W. (2009). If Someone Is Watching, I'll Do what I'm Asked: Mandatoriness, Control, and Information Security. *Eur. J. Inf. Syst.* 18 (2), 151–164. doi:10.1057/ejis.2009.8

Brett, J. F., and VandeWalle, D. (1999). Goal Orientation and Goal Content as Predictors of Performance in a Training Program. *J. Appl. Psychol.* 84 (6), 863–873. doi:10.1037/0021-9010.84.6.863

Burns, A. J., Roberts, T. L., Posey, C., and Lowry, P. B. (2019). The Adaptive Roles of Positive and Negative Emotions in Organizational Insiders' Security-Based Precaution Taking. *Inf. Syst. Res.* 30 (4), 1228–1247. doi:10.1287/isre.2019.0860

Canham, M., Posey, C., Strickland, D., and Constantino, M. (2021). Phishing for Long Tails: Examining Organizational Repeat Clickers and Protective Stewards. *SAGE Open* 11 (1), 2158244021990656. doi:10.1177/2158244021990656

Caputo, D. D., Pfleeger, S. L., Freeman, J. D., and Johnson, M. E. (2013). Going Spear Phishing: Exploring Embedded Training and Awareness. *IEEE Security & Privacy* 12 (1), 28–38.

Cerasoli, C. P., and Ford, M. T. (2014). Intrinsic Motivation, Performance, and the Mediating Role of Mastery Goal Orientation: A Test of Self-Determination Theory. *J. Psychol.* 148 (3), 267–286. doi:10.1080/00223980.2013.783778

CheckPoint (2021). Biggest Cyber Security Challenges in 2021. Retrieved from https://www.checkpoint.com/cyber-hub/cyber-security/what-is-cybersecurity/biggest-cyber-security-challenges-in-2021/.

Conley, C. (2021). Ethical Phishing –The Slippery Slope with Employee Deception. Retrieved from https://www.sans.org/blog/ethical-phishing-the-slippery-slope-with-employee-deception/.

Dincelli, E., and Chengalur-Smith, I. (2020). Choose Your Own Training Adventure: Designing a Gamified SETA Artefact for Improving Information Security and Privacy through Interactive Storytelling. *Eur. J. Inf. Syst.* 29 (6), 669–687. doi:10.1080/0960085x.2020.1797546

Emm, D. (2021). Gamification - Can it Be Applied to Security Awareness Training? *Netw. Security* 2021 (4), 16–18. doi:10.1016/s1353-4858(21)00040-4

Ferrell, S. (2021). The Problem with Phishing Simulators. Retrieved from https://www.inky.com/blog/the-problem-with-phishing-simulators.

Fleming, T., Sutcliffe, K., Lucassen, M., Pine, R., and Donkin, L. (2020). "Serious Games and Gamification in Clinical Psychology," in *Reference Module in Neuroscience and Bio Behavioral Psychology* (Elsevier). doi:10.1016/b978-0-12-818697-8.00011-x

Francia, G., III, Thornton, D., Trifas, M., and Bowden, T. (2014). "Gamification of Information Security Awareness Training," in *Emerging Trends in ICT Security* (Elsevier), 85–97. doi:10.1016/b978-0-12-411474-6.00005-0

Gartner (2021). Gartner Forecasts Worldwide Security and Risk Management Spending to Exceed $150 Billion in 2021. Retrieved from https://www.gartner.com/en/newsroom/press-releases/2021-05-17-gartner-forecasts-worldwide-security-and-risk-managem.

Gjertsen, E. G. B., Gjære, E. A., Bartnes, M., and Flores, W. R. (2017). "Gamification of Information Security Awareness and Training," in Paper Presented at the ICISSP (Setúbal, Portugal: SciTePress). doi:10.5220/0006128500590070

Gong, Y., and Fan, J. (2006). Longitudinal Examination of the Role of Goal Orientation in Cross-Cultural Adjustment. *J. Appl. Psychol.* 91 (1), 176–184. doi:10.1037/0021-9010.91.1.176

Greene, K. K., Steves, M., Theofanos, M., and Kostick, J. (2018). "User Context: an Explanatory Variable in Phishing Susceptibility," in Paper Presented at the in Proc. 2018 Workshop Usable Security (NY, United States: Association for Computing Machinery). doi:10.14722/usec.2018.23016

Groening, C., and Binnewies, C. (2019). "Achievement Unlocked!" - the Impact of Digital Achievements as a Gamification Element on Motivation and Performance. *Comput. Hum. Behav.* 97, 151–166. doi:10.1016/j.chb.2019.02.026

Halevi, T., Memon, N., and Nov, O. (2015). *Spear-phishing in the Wild: A Real-World Study of Personality, Phishing Self-Efficacy and Vulnerability to Spear-Phishing Attacks*. Rochester, NY: SSRN.

Kaplan, A., and Maehr, M. L. (2007). The Contributions and Prospects of Goal Orientation Theory. *Educ. Psychol. Rev.* 19 (2), 141–184. doi:10.1007/s10648-006-9012-5

Karac, J., and Stabauer, M. (2017). "Gamification in E-Commerce-A Survey Based on the Octalysis Framework," in International Conference on HCI in Business, Government, and Organizations (Springer), 41–54.

Khando, K., Gao, S., Islam, S. M., and Salman, A. (2021). Enhancing Employees Information Security Awareness in Private and Public Organisations: A Systematic Literature Review. *Comput. Security* 106, 102267. doi:10.1016/j.cose.2021.102267

Khonji, M., Iraqi, Y., and Jones, A. (2013). Phishing Detection: a Literature Survey. *IEEE Commun. Surv. Tutorials* 15 (4), 2091–2121. doi:10.1109/surv.2013.032213.00009

Kumaraguru, P., Rhee, Y., Acquisti, A., Cranor, L. F., Hong, J., and Nunge, E. (2007). "Protecting People from Phishing: the Design and Evaluation of an Embedded Training Email System," in Paper Presented at the Proceedings of the SIGCHI Conference on Human Factors in

Computing Systems (NY, United States: Association for Computing Machinery).

Lawson, P., Zielinska, O., Pearson, C., and Mayhorn, C. B. (2017). "Interaction of Personality and Persuasion Tactics in Email Phishing Attacks," in Paper Presented at the Proceedings of the Human Factors and Ergonomics Society Annual Meeting (Santa Monica, CA: Human Factors and Ergonomics Society). doi:10.1177/1541931213601815

Lewis, Z. H., Swartz, M. C., and Lyons, E. J. (2016). What's the point?: a Review of Reward Systems Implemented in Gamification Interventions. *Games Health J.* 5 (2), 93–99. doi:10.1089/g4h.2015.0078

Maples-Keller, J. L., Williamson, R. L., Sleep, C. E., Carter, N. T., Campbell, W. K., and Miller, J. D. (2019). Using Item Response Theory to Develop a 60-Item Representation of the NEO PI-R Using the International Personality Item Pool: Development of the IPIP-NEO-60. *J. Pers Assess.* 101 (1), 4–15. doi:10.1080/00223891.2017.1381968

Marín, B., Frez, J., Cruz-Lemus, J., and Genero, M. (2018). An Empirical Investigation on the Benefits of Gamification in Programming Courses. *ACM Trans. Comput. Edu. (Toce)* 19 (1), 1–22.

McCormac, A., Zwaans, T., Parsons, K., Calic, D., Butavicius, M., and Pattinson, M. (2017). Individual Differences and Information Security Awareness. *Comput. Hum. Behav.* 69, 151–156. doi:10.1016/j.chb.2016.11.065

McCrae, R. R., and Costa, P. T. (1987). Validation of the Five-Factor Model of Personality across Instruments and Observers. *J. Pers Soc. Psychol.* 52 (1), 81–90. doi:10.1037//0022-3514.52.1.81

Meixner, C., Baumann, H., and Wollesen, B. (2020). Personality Traits, Gamification and Features to Develop an App to Reduce Physical Inactivity. *Information* 11 (7), 367. doi:10.3390/info11070367

Norman, W. T. (1963). Toward an Adequate Taxonomy of Personality Attributes: Replicated Factors Structure in Peer Nomination Personality Ratings. *J. Abnorm Soc. Psychol.* 66 (6), 574–583. doi:10.1037/h0040291

Pattinson, M., Jerram, C., Parsons, K., McCormac, A., and Butavicius, M. (2012). *Why Do Some People Manage Phishing E-mails Better than Others?* Bingley, United Kingdom: Information Management & Computer Security.

Payne, S. C., Youngcourt, S. S., and Beaubien, J. M. (2007). A Meta-Analytic Examination of the Goal Orientation Nomological Net. *J. Appl. Psychol.* 92 (1), 128–150. doi:10.1037/0021-9010.92.1.128

Porter, C. O., Webb, J. W., and Gogus, C. I. (2010). When Goal Orientations Collide: Effects of Learning and Performance Orientation on Team Adaptability in Response to Workload Imbalance. *J. Appl. Psychol.* 95 (5), 935–943. doi:10.1037/a0019637

Posey, C., Roberts, T. L., Roberts, T. L., Lowry, P. B., Bennett, R. J., and Courtney, J. F. (2013). Insiders' protection of Organizational Information Assets: Development of a Systematics-Based Taxonomy and Theory of Diversity for protection-motivated Behaviors. *Misq* 37, 1189–1210. doi:10.25300/misq/2013/37.4.09

Purplesec (2021). 2021 Cyber Security Statistics: The Ultimate List of Stats, Data & Trends. Retrieved from https://purplesec.us/resources/cyber-security-statistics/.

Schneier, B. (2015). *Secrets and Lies: Digital Security in a Networked World*. John Wiley & Sons.

Scholefield, S., and Shepherd, L. A. (2019). "Gamification Techniques for Raising Cyber Security Awareness," in Paper Presented at the International Conference on Human-Computer Interaction (Springer). doi:10.1007/978-3-030-22351-9_13

Seligman, M. E. P., and Csikszentmihalyi, M. (2014). "Positive Psychology: An Introduction," in *Flow and the Foundations of Positive Psychology* (Springer), 279–298. doi:10.1007/978-94-017-9088-8_18

Shappie, A. T., Dawson, C. A., and Debb, S. M. (2020). Personality as a Predictor of Cybersecurity Behavior. *Psychol. Popular Media* 9 (4), 475–480. doi:10.1037/ppm0000247

Silic, M., and Lowry, P. B. (2020). Using Design-Science Based Gamification to Improve Organizational Security Training and Compliance. *J. Manag. Inf. Syst.* 37 (1), 129–161. doi:10.1080/07421222.2019.1705512

Steves, M., Greene, K., and Theofanos, M. (2020). Categorizing Human Phishing Difficulty: a Phish Scale. *J. Cybersecurity* 6 (1), tyaa009. doi:10.1093/cybsec/tyaa009

Straub, D. W., and Nance, W. D. (1990). Discovering and Disciplining Computer Abuse in Organizations: A Field Study. *MIS Q.* 14, 45–60. doi:10.2307/249307

Sudzina, F., and Pavlicek, A. (2017). *Propensity to Click on Suspicious Links: Impact of Gender, of Age, and of Personality Traits*. Bled, Austria: BLED.

Uebelacker, S., and Quiel, S. (2014). "The Social Engineering Personality Framework," in Paper Presented at the 2014 Workshop on Socio-Technical Aspects in Security and Trust (Vienna, Austria: IEEE). doi:10.1109/stast.2014.12

Ueyama, Y., Tamai, M., Arakawa, Y., and Yasumoto, K. (2014). "Gamification-based Incentive Mechanism for Participatory Sensing," in 2014 IEEE International Conference on Pervasive Computing and Communication Workshops (PERCOM WORKSHOPS) (IEEE), 98–103. doi:10.1109/percomw.2014.6815172

Wash, R., and Cooper, M. M. (2018). "Who Provides Phishing Training? Facts, Stories, and People like Me," in Paper Presented at the Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (New York: ACM).

Welk, A. K., Hong, K. W., Zielinska, O. A., Tembe, R., Murphy-Hill, E., and Mayhorn, C. B. (2015). Will the "Phisher-Men" Reel You in? *Int. J. Cyber Behav. Psychol. Learn. (Ijcbpl)* 5 (4), 1–17. doi:10.4018/ijcbpl.2015100101

Willison, R., Warkentin, M., and Warkentin, M. (2013). Beyond Deterrence: An Expanded View of Employee Computer Abuse. *Misq* 37, 1–20. doi:10.25300/misq/2013/37.1.01

Workman, M., Bommer, W. H., and Straub, D. (2008). Security Lapses and the Omission of Information Security Measures: A Threat Control Model and Empirical Test. *Comput. Hum. Behav.* 24 (6), 2799–2816. doi:10.1016/j.chb.2008.04.005

Zielinska, O. A., Tembe, R., Hong, K. W., Ge, X., Murphy-Hill, E., and Mayhorn, C. B. (2014). "One Phish, Two Phish, How to Avoid the Internet Phish: Analysis of Training Strategies to Detect Phishing Emails," in Paper Presented at the Proceedings of the Human Factors and Ergonomics Society Annual Meeting (Santa Monica, CA: Human Factors and Ergonomics Society).

Conflict of Interest: Author MCa was employed by the company Beyond Layer Seven, LLC.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.