Comparing Phishing Training and Campaign Methods for Mitigating Malicious Emails in Organizations

Jackie (Chris) Scott[a], Yair Levy[b], Wei Li[c] and Ajoy Kumar[d]
College of Computing and Engineering, Nova Southeastern University,
3301 College Avenue, Fort Lauderdale, Florida, 33314, U.S.A.

Keywords: Phishing, Spear-Phishing, Security Education, Training, and Awareness (SETA), Business Email Compromise (BEC), Red Team, Phishing Campaigns, Phishing Training.

Abstract: Although there have been numerous significant technological advancements in the last several decades, there continues to be a real threat as it pertains to social engineering, especially phishing, spear-phishing, and Business Email Compromise (BEC). While the technologies to protect end-users have gotten better, the 'human factor' in cybersecurity is the main penetration surface. These three phishing methods are used by attackers to infiltrate corporate networks and manipulate end-users, especially through business email. Our research study was aimed at assessing several phishing mitigation methods, including phishing training and campaign methods, as well as any human characteristics that enable a successful cyberattack through business email. Following expert panel validation for the experimental procedure, a pilot study with 172 users and then a full study with 552 users were conducted to collect six actual end-users' negative response actions to phishing campaigns conducted with traditional Commercial-Off-The-Shelf (COTS) product (KnowBe4) and a red team. Users were randomly assigned to three groups: no training; traditional training; and longitudinal customized training with 1,104 data points collected. While the phishing method was significant, our results indicate that current training methods appear to provide little to no added value vs. no training at all.

1 INTRODUCTION

The role of Information Security (ISec) continues to be the first line of defense in guarding Personally Identifiable Information (PII) of the end-users of modern Information Systems (IS) (Ho, 2018). As the focus on email threats continues to strengthen from attackers, it is imperative to research the success factors of these attacks so that steps can be taken to guard against email sabotage, financial ransom, email compromise, and hacking (Costantino et al., 2018). Salahdine and Kaabouch (2019) stated that social engineering, specifically phishing and Business Email Compromise (BEC) campaigns, are on the rise and it is critical to understand the factors behind it and the impact on Intellectual Property (IP). While the research on social engineering goes back to the mid-1990s, the research on specific phishing and BEC topics, outcomes, and mitigations are starting to gain momentum in academia as it is a serious threat to IP.

The research problem that this study addressed is the continued growth of cyberattacks targeting businesses via email to impersonate the corporate end-user for causing significant financial damages to organizations (FBI, 2019). Further, lack of cybersecurity knowledge and skills contribute to the enablement of up to 95% of cybersecurity threats, which lead to significant financial and IP loss to businesses (Carlton & Levy, 2015). The Federal Bureau of Investigation (FBI) Internet Crime Complaint Center (IC3) stated that successful cybercrime attacks were responsible for $4.2 billion in financial losses in 2020, in the United States (U.S.) alone (FBI IC3, 2020). Phishing is one type of social engineering, which is defined as a scalable act of deception whereby impersonation is used to obtain

[a] https://orcid.org/0009-0002-2671-6982
[b] https://orcid.org/0000-0002-8994-6497
[c] https://orcid.org/0000-0001-5880-4640
[d] https://orcid.org/0000-0001-6450-7730

Scott, J., Levy, Y., Li, W. and Kumar, A.
Comparing Phishing Training and Campaign Methods for Mitigating Malicious Emails in Organizations.
DOI: 10.5220/0012281600003648
Paper published under CC license (CC BY-NC-ND 4.0)
In *Proceedings of the 10th International Conference on Information Systems Security and Privacy (ICISSP 2024)*, pages 643-651
ISBN: 978-989-758-683-5; ISSN: 2184-4356

information from the target (Lastdrager, 2014). These attacks have become increasingly more sophisticated with attackers using very customized business emails to lure corporate end-users to trust and act on them (Kotson & Shultz, 2015).

The main goal of our research study was to compare email phishing training methods (annual industry-standard awareness training and continuous customized social engineering Security Education, Training, and Awareness (SETA) program) and email phishing campaign methods (industry-standard phishing campaign and a Red Team phishing campaign) and their role in mitigating simulated email phishing attempts in organizations. The traditional Commercial-Off-The-Shelf (COTS) platform utilized to deliver both the SETA and traditional phishing campaigns was KnowBe4 (https://www.knowbe4.com/). KnowBe4 was used to create one of the email phishing campaigns, and the other was created by a Red Team during the penetration testing period. KnowBe4 also acted as the single instrument for gathering data on the success of malicious emails when delivered to corporate end-users. The secondary goal of our research study was to assess any statistical mean differences between the three groups of users (two types of training and no training), the campaign methods (via COTS or via Red Team), and their impact on actual email phishing mitigation behavior by the corporate end-users when controlled by five demographic factors.

1.1 Research Questions

RQ1: What are the approved components of the experimental procedures for the phishing training and campaign methods according to cybersecurity Subject Matter Experts (SMEs)?
RQ2: What level of validity of the experimental procedures the phishing training and campaign methods is sufficient according to cybersecurity SMEs?
RQ3: Are there any statistically significant mean differences between the use of an annual industry-standard phishing training, continuous customized social engineering focused training, and a control group without training, on end-users' negative response to malicious emails?
RQ4: Are there any statistically significant mean differences between the use of an industry-standard phishing campaign and a Red Team phishing campaign on end-users' negative response to malicious emails?
RQ5: Are there any statistically significant mean differences between the phishing training methods (an annual industry-standard phishing awareness training vs. continuous customized social engineering-focused training vs. no training - control) and the phishing campaign methods (industry-standard phishing campaign vs. Red Team phishing campaign) on end-users' negative response to malicious emails?
RQ6: Are there any statistically significant mean differences between the use of an annual industry-standard phishing training, continuous customized social engineering-focused training, and a control group without training, on end-users' negative response to malicious emails, when controlled for participants': (a) age, (b) gender, (c) job role, (d) location (clinic vs. admin), and (e) years of job experience?
RQ7: Are there any statistically significant mean differences between the use of an industry-standard phishing campaign and a Red Team phishing campaign on end-users' negative response to malicious emails, when controlled for participants': (a) age, (b) gender, (c) job role, (d) location (clinic vs. admin), and (e) years of job experience?

2 BACKGROUND

A literature review was conducted to provide a theoretical foundation for our research study focused on phishing mitigation methods. Based on the overall increase in cyberattacks, and the many high-profile ransomware and data theft cases over the last several years, phishing mitigation remains a primary goal in every organization. Most current statistics state that over 90% of successful cyberattacks and breaches begin as a phish. With every organization's dependency on email, coupled with the fact that over three billion phishing emails are sent every day around the globe (Earthweb, 2022), it appears there is a need to continue researching phishing mitigation methods.

A literature review of social engineering, SETA, and email phishing mitigation methods in the cybersecurity research field has been conducted to provide a foundation for this research study. While many studies have focused on types of social engineering, phishing continues to be the number one cause of successful cyberattacks in organizations. The FBI IC3 (2021) stated over 90% of all data breaches start as a phish and may be increasing by as much as 400% per year. Phishing success and its negative impact on organizations is widely publicized and known, however, this research study will focus on what is unknown in this field. What is unknown, as it relates to successful phishing email attacks in organizations, is the impact that various forms of phishing training programs coupled with various phishing campaign methods have on phishing mitigation. Another mitigation method, the Intrusion Detection and Prevention System (IDPS), was also discussed as part of the literature review, however, the focus of our study was contained within the corporate email environment and not the external network. We believe that our study addresses a current gap in the body of knowledge as it relates to phishing attacks and how to effectively mitigate them in an organizational environment.

3 METHODOLOGY

This experimental research targeted the difference between phishing training methods and phishing campaign methods when controlled by multiple factors. In Phase 1, we developed a baseline measure between training and campaign results leveraging an expert panel of cybersecurity professionals utilizing the Delphi method. The expert panel consisted of 50 cybersecurity SMEs to conduct the review. The Delphi method is a demonstrated technique in the field of IS in the development of the experiment with SMEs (Ramim & Lichvar, 2014). Phase 2 of our study included a randomized participant sample selection of 30 business professionals for each quasi-experiment. Phase 3 of our study further expanded on phishing training methods versus phishing campaign methods results but was controlled by demographic indicators and vulnerability action types to quantify any statistically significant differences. The specific end-user negative response actions measured during Phase 2 and Phase 3 are noted in Table 1.

Table 1: The Six End-User Negative Response Actions Measures (Increase in Severity).

No.	End-User Action
1.	Non-identification of phishing email
2.	Clicking/opening a phishing email
3.	Replying/forwarding to others
4.	Opening/executing attachments
5.	Enabling macros
6.	Data entry to a malicious website

3.1 SMEs Instrument

The 50 targeted cybersecurity SMEs were recruited through many different methods, including social media (Facebook, LinkedIn, etc.), as well as word of mouth and personal network facilitation. Once the SME panel was finalized, each person received a link to the "Cybersecurity SME Survey" using the Google Forms® platform. Ultimately this survey confirmed the approved components of the experimental procedures, as well as validated their use as part of the research experiments. The outcome of the SME instrument results was used to positively confirm RQ1 and RQ2 of this research study.

3.2 Organizational End-User Instrument

This study leveraged a COTS platform that provided reporting as to the behavior of the end-user. During the email phishing campaigns, the negative response actions related to the vulnerability types were logged for every email and every end-user in that specific campaign. This industry-standard reporting platform (KnowBe4) provided a detailed analysis of the campaign, regardless of method (Industry-standard or Red Team).

3.3 Data Analysis

3.3.1 Phase I: SMEs Validation

Quantitative data collection methods were used in Phase 1 for the collection of cybersecurity SMEs' inputs with validation of current email phishing mitigation methods, as well as email phishing training and campaign methods. The specific data collection method was a short survey sent by email to the selected SMEs. According to Kost and Correa da Rosa (2018), a shorter survey instrument holds the potential to dramatically improve the response rate as opposed to a longer survey. This shorter survey was created utilizing a 7-point Likert scale for non-demographic questions and will rate agreement from (1) Strongly Disagree through (7) Strongly Agree.

The Delphi methodology was used to ensure the reliability and validity of the instrument utilized for this research study. This methodology is oftentimes used to summarize the agreement between the SME group as to the applicability of the measurement instrument. Walker and Selfe (1996) stated that a 70% agreement in the survey questions by respondents was an acceptable rate to validate the instrument and move forward with this study, which we followed.

3.3.2 Phase II: Pilot Study

Phase 2 consisted of a pilot study with randomized participants grouped into one of two developed treatments (Industry-standard and Red Team) as well as a control group (no training). Pilot data was collected, and data analysis was performed using one-way Analysis of Variance (ANOVA). The experiment was revised per the preliminary data analysis and the results aided in adjusting the research measures to ensure internal validity. This study utilized the linear statistical models to address the research questions utilizing SPSS® Statistics™ version 28. The statistical analysis one-way ANOVA was used to assess significant mean differences between the variables being studied (Sekaran & Bougie, 2016).

3.3.3 Phase III: Main Study

Phase 3 incorporated the findings from the pilot study in Phase 2 and used this information to perform the main study. All data gathered on the population came from a System Administrator on the Chief Technology Officers (CTO) team. In addition, all needed information to codify and analyze the data was provided by this separate team. All data to answer demographic questions as part of RQ6 and RQ7 was provided by the team based on the employee ID of the participants. All end-user participants were required to provide consent in email to be considered for the research study. No PII was provided during data collection for the experiments per IRB guidelines.

4 RESULTS

4.1 Phase I: Cybersecurity SME Survey Feedback

RQ1 and RQ2 were answered through a survey instrument during the first phase of this research study. Participation in the Cybersecurity SME survey was facilitated by sending an email invitation to 50 potential candidates within the network of work, school, and personal acquaintances, with a goal of 25 respondents. Of the 50 potential candidates, 27 or 54% response rate, cybersecurity SMEs completed the survey over a period of about three weeks.

4.1.1 Phase I: SMEs Validation

In addition to the demographics captured, the cybersecurity SME group also provided inputs to answer both RQ1 and RQ2. The survey answers provided positive feedback on both the approved components and the level of validity of the experiment based on Delphi consensus thresholds. In general, Delphi consensus thresholds range from 51% to 100%, however, a 75% or greater score is standard and, therefore, is an acceptable threshold for decision-making (Dupuis et al., 2016). For RQ1, the SME panel was asked to rate the six "end-user negative response actions" used to score the experiment (Table 2), as well as the two different campaign methods that were to be used in both the pilot and main studies (Table 3).

Table 2: SME % Agreement for Six End-user Negative Response Actions (N=27).

	User not able to identify	Clicking / Opening	Replying / Forwarding	Opening attachment	Enabling a macro	Data entry when prompted
Averages	5.93	6.04	5.74	6.52	6.11	6.30
Stand Dev	1.2066	1.2855	1.5589	0.9352	1.6718	1.1373
% Agreement	85%	89%	85%	96%	85%	93%

Table 3: SME % Agreement for Phishing Campaign Methods (N=27).

	Industry-standard	Red Team
Averages	5.44	5.85
Stand Dev	1.7172	1.6572
% Agreement	81%	89%

In addition to the approval of the experimental procedures (RQ1), the SME respondents were asked to provide feedback on the validity of the overall experiment and the significance of phishing overall in the world of cybersecurity (RQ2). Based on the responses regarding the SMEs level of knowledge around phishing and phishing campaigns (96% agreement) coupled with the direct question about the "significance of phishing today" (also 96% agreement) showed clear support of the measures..

4.2 Phase II: Pilot Study

4.2.1 Pilot Study Data Collection

The pilot study was conducted to confirm the ability to acquire the needed data for the main study, as well as to test the procedures in which the data was collected. In this phase, a random selection of three distinct groups of organizational end-users were chosen to test and confirm the process. First, a group of 30 unique organizational end-users were selected who were new to the organization and had not been exposed to any previous phishing training from the company. Second, a separate group of 30 unique organizational end-users were selected to receive an annual, industry-standard phishing training of 30 minutes. Last, a separate group of 30 unique organizational end-users were selected to receive customized, continuous phishing training, consisting of eight short videos of less than five minutes in length.

After the 4-week training phase of the data collection process, the phishing campaigns commenced. Each week for two weeks, all three groups were phished twice a week by both phishing campaign methods (Industry-standard and Red Team). The KnowBe4 platform was able to collect data on all six negative end-user responses by providing a count of clicks for each action. The pilot data upon completion consisted of two sets of 86 responses (No training 28, Annual 29, Continuous 29) for 172 discrete responses to analyze. During the pilot data collection phase, there were no demographic indicators captured. To further confirm the data collection process was accurate, an analysis of the pilot data was completed to test the results for RQ3 and RQ4.

4.2.2 Pilot Data Analysis

Table 4 shows the output of the one-way ANOVA for RQ3 to determine any mean differences.

Table 4: One-way ANOVA Output for RQ3 Using Pilot Data (N=172).

Negative Response Action	**Sum of Squares**	**df**	**Mean Square**	**F**	**Sig.**
Not Reported	.822	2	.411	1.151	.319
Opened	.166	2	.083	.109	.897
Reply/Forward	.012	2	.006	1.036	.357
Open Attachment	.000	2	.000	N/A	N/A
Enabled Macro	2.080	2	1.040	1.766	.174
Entered Data	1.825	2	.913	.746	.476

* *p* <0.05, ** *p* <0.01, *** *p* <0.001

Based on the output of the ANOVA, there appear to be no statistically significant mean differences between the phishing training method and the six negative response actions (p values above 0.05, see Table 4). This result would indicate that the training method is overall not an important factor in determining negative end-user response actions. Table 5 shows the output of the one-way ANOVA for RQ4 to determine any mean differences.

Based on the output of the ANOVA, there appear to be several statistically significant mean differences between the phishing campaign method and the six negative end-user response actions. The mean differences for four of the six were statistically significant (See Table 5). This result indicates the way the phishing campaign method is conducted has a significant impact on end-user negative response actions.

Table 5: One-way ANOVA Output for RQ4 Using Pilot Data (N=172)

Negative Response Action	**Sum of Squares**	**df**	**Mean Square**	**F**	**Sig.**
Not Reported	2.814	1	2.814	8.195	***.005****
Opened	11.256	1	11.256	16.146	***<.001*****
Reply/Forward	.006	1	.006	1.000	.319
Open Attachment	.000	1	.000	N/A	N/A
Enabled Macro	3.930	1	3.930	6.840	***.010****
Entered Data	7.535	1	7.535	6.375	***.012****

* *p* <0.05, ** *p* <0.01, *** *p* <0.001

4.3 Phase III: Main Study

4.3.1 Main Study Data Collection

Like the pilot, in this phase, a random selection of three distinct groups of organizational end-users was chosen to create the main study dataset. First, a group of 200 unique organizational end-users were selected who were new to the organization and had not been exposed to any previous phishing training from the company. Second, a separate group of 200 unique organizational end-users were selected to receive an annual, industry-standard phishing training of 30 minutes. Last, a separate group of 200 unique organizational end-users were selected to receive customized, continuous phishing training, consisting of eight short videos of less than five minutes in length.

This initial recruitment process lasted for one week to provide the organization end-user an opportunity to respond. The results for the main study groups showed only a 7-9% opt-out rate. For the no-training group of 200, 17 users responded no to participating in this study. For the annual training group of 200, 15 users responded no to participating in this study, and for the continuous customized group of 200, 16 users declined. The main study data upon completion consisted of two sets of 552 responses (No training 183, Annual 185, Continuous 184) for 1,104 discrete responses to analyze.

4.3.2 Main Study Data Analysis

Table 6 shows the output of the one-way ANOVA for RQ3 to determine any mean differences.

Based on the ANOVA results there is one statistically significant mean difference between the phishing training method and one of the six negative end-user response actions (Opened (F (2,1103) = 4.722, p = .009). Using the Tukey HSD output for

Table 6: One-way ANOVA Output for RQ3 (N=1104).

Negative Response Action	Sum of Squares	df	Mean Square	F	Sig.
Not Reported	2.270	2	1.135	1.266	.282
Opened	19.211	2	9.605	4.722	***.009***
Reply/Forward	.002	2	.001	.121	.886
Open Attachment	.000	2	.000	N/A	N/A
Enabled Macro	1.516	2	.758	2.697	.068
Entered Data	1.770	2	.885	1.180	.308

* *p* <0.05, ** *p* <0.01, *** *p* <0.001

multiple comparisons (Table 7), there is a statistically significant difference between training Group 2 (Annual) and training Group 3 (Continuous Customized) as it relates to an end-user opening a phishing email (p < 0.5, see Table 5).

Table 7: Tukey HSD Output for RQ3 (Opened).

Dep Var	(I)Train Grp	(J)Train Grp	Mean Diff (I-J)	Std. Error	Sig.	Lower Bound	Upper Bound
Opened	1	2	-.244	.105	.053	-.49	.00
		3	.061	.105	.832	-.19	.31
	2	1	.244	.105	.053	.00	.49
		3	.305	.105	***.011****	.06	.55
	3	1	-.061	.105	.832	-.31	.19
		2	-.305	.105	***.011****	-.55	-.06

* *p* <0.05, ** *p* <0.01, *** *p* <0.001

Table 8 shows the output of the one-way ANOVA for RQ4 to determine any mean differences.

Table 8: One-way ANOVA Output for RQ4 (N=1104).

Negative Response Action	Sum of Squares	df	Mean Square	F	Sig.
Not Reported	208.696	1	208.696	294.620	***<.001****
Opened	315.308	1	315.308	178.780	***<.001****
Reply/Forward	.058	1	.058	8.103	***.005***
Open Attachment	.000	1	.000	N/A	N/A
Enabled Macro	.110	1	.110	.389	.533
Entered Data	21.204	1	21.204	28.970	***<.001****

* *p* <0.05, ** *p* <0.01, *** *p* <0.001

Based on the output of the ANOVA, there appears to be several statistically significant mean differences between phishing campaign method and the six negative end-user response actions. The mean differences for Not Reported (F (1,1103) = 294.620, p = < .001), Opened (F (1,1103) = 178.780, p = < .001), Reply/Forward (F (1,1103) = 8.103, p = .005), and Entered Data (F (1,1103) = 28.970, p = < .001) are all statistically significant. This result indicates the way the phishing campaign method is delivered (Industry-standard vs. Red Team) has a significant impact on end-user negative response actions. Both pilot and main study data results for RQ4 are consistent and statistically significant.

Table 9: ANCOVA Output for RQ5 (N=1104).

Method	Type III Sum of Squares	df	Mean Square	F	Sig.
Training Group	.688	1	.688	.390	.532
Campaign Method	315.308	1	315.308	178.681	***<.001****

* *p* <0.05, ** *p* <0.01, *** *p* <0.001

Table 9 shows the output of the Analysis of Covariance (ANCOVA) for RQ5 to determine any mean differences. Based on the output of the Tests of Between-Subjects Effects ANCOVA, there appears to be a statistically significant mean difference between the phishing training group and the phishing campaign. The mean differences for Campaign Method (F (1,1103) = 178.681, p = < .001) show that overall, the campaign method is most important as it relates to phishing success for the six negative end-user response actions measured. To take this a step further, all the data from each phishing campaign was summarized to show click rates by each training method. Table 10 shows that across all three training groups, the Red Team campaign method was the most successful in getting end-users to take a negative action.

Table 10: Average Click Rates Across Training Groups.

# Users	Oppts	Total Oppts	# Clicks	Click %	Train Grp	Campaign Type
183	4	732	213	29.10	No Train	Industry Std
183	4	732	268	36.61	No Train	Red Team
185	4	740	263	35.54	Annual	Industry Std
185	4	740	419	56.62	Annual	Red Team
184	4	736	225	30.57	Continuous	Industry Std
184	4	736	274	37.23	Continuous	Red Team
				31.74	Avg %	Industry Std
				43.49	Avg %	Red Team

Given the statistically significant findings of the ANCOVA, and the average click rates across training groups, the use of the Red Team method for phishing campaigns appears more effective than Industry-standard. In addition, it seems this holds true no matter the type of training method the end-user receives.

Table 11: ANCOVA Output for RQ6 with Demographic Control (N=1104).

Demographic Variable	Type III Sum of Squares	df	Mean Square	F	Sig.
Age	.049	1	.049	.074	.785
Gender	2.026	1	2.026	3.045	.081
Job Role	1.789	1	1.789	2.689	.101
Location	.007	1	.007	.011	.917
YoE	1.169	1	1.169	1.757	.185

* p <0.05, ** p <0.01, *** p <0.001

To answer RQ6 and RQ7 there were five demographic indicators added to the main study dataset. The demographic indicators the final two RQs were controlled for are: (a) age, (b) gender, (c) job role, (d) location (clinic vs. admin), and (e) years of job experience. Table 11 shows the output of the ANCOVA for RQ6 to determine any mean differences. Given the results of the ANCOVA there appears no statistically significant mean differences for the phishing training group on the six negative end-user response actions when controlled for demographic indicators. Table 12 shows the output of the ANCOVA for RQ7 to determine any mean differences.

Table 12: ANCOVA Output for RQ7 with Demographic Control (N=1104).

Demographic Variable	Type III Sum of Squares	df	Mean Square	F	Sig.
Age	.056	1	.056	.429	.513
Gender	.009	1	.009	.069	.793
Job Role	.116	1	.116	.900	.343
Location	.008	1	.008	.063	.802
YoE	.001	1	.001	.004	.948

* p <0.05, ** p <0.01, *** p <0.001

Given the results of the ANCOVA there appears no statistically significant mean differences for phishing campaign method on the six negative end-user response actions when controlled for demographic indicators.

4.4 Summary of Results

Phase 1 of this research study included utilizing cybersecurity SMEs via the Delphi process to confirm the approved components and the level of validity of the research study. A consensus was reached on all measurements (six negative end-user response actions) and methods, and this study was approved to move forward. The cybersecurity SME survey results were used to answer RQ1 and RQ2.

In Phase 2, the pilot study, a 7-week process for randomly selecting organizational end-user training groups was created. Following this was a 4-week training cycle for each of the three phishing training groups (No training, Annual training, and Continuous Customized training). The last part of the process consisted of a 2-week phishing campaign, in which each participant was phished twice a week by both phishing campaign methods (Industry-standard and Red Team). The result of the 7-week process was a clean and accurate dataset produced by the KnowBe4 platform to analyze.

In Phase 3, the main study, the same 7-week process was utilized to randomly select a larger set of organizational end-users to participate. With the pilot study, three training groups of 30 were defined, however, the main study was significantly larger by utilizing three groups of 200. The result at the end of the data collection process was 1,104 unique data points, which were used to formally answer RQ3, RQ4, and RQ5. For the main study, demographic indicators were also attached to the organizational end-user record to provide answers to RQ6 and RQ7.

5 CONCLUSIONS

The findings of our research study are significant as they contribute to the body of knowledge and have several key implications for providing both researchers and practitioners additional insight into mitigating phishing attacks. The indication that phishing training methods have little effect overall on end-user negative response actions should imply that new ways of dealing with phishing from the technological perspective should be developed, without the need for end-user intervention. According to our results, annual cybersecurity awareness training is not effective in mitigating corporate networks from phishing threats. Annual training, and even continuous customized training, are still delivered in video format and are easily dismissed. Despite indications that an end-user has successfully completed a module or video really has little meaning today. An implication from this study should be to rethink how organizational end-users are trained and find a new dynamic approach that is more efficient and effective. SETA is critical to ensure the user population is aware of the risks, but modernization of the approach and technological methods are imperative. The indication that phishing campaign methods are statistically significant should imply that organizations must continue phishing campaigns, but also learn from the results and act. Our study indicates that a vended Red Team campaign was most effective in phishing the end-user population. While not every organization is large enough to support an internal Red Team, it is imperative this method is utilized, even on a contract or third-party vendor basis.

5.1 Recommendations and Future Research

This research study was to compare phishing training and campaign methods and their role in mitigating malicious emails in organizations. While the goals of our research study were met, there are many areas for expansion and additional future research in the phishing training and campaign method domains. The implications above also lead to recommendations on how to continuously improve this process. As stated, the IT security industry needs to rethink current ways of training end-users, and their overall effectiveness. With many current advances in Machine Learning (ML) and Artificial Intelligence (AI) there stands to be a great opportunity to address this issue. By combining ML/AI methods with modern advances in behavioral technology, there must be a better way to prevent, not just mitigate, phishing attacks. Future research in this area should include more focus on technological approaches to prevent phishing emails by providing alerts and warnings similar to not fastening a car seatbelt (Cooper et al., 2021) and complete screen freezing techniques to enable the end-users to shift from System 1 to System 2 thinking (Antonucci et al., 2022). This research study was conducted in a medium-sized, privately held, healthcare network organization. The composition of the organization is typical for a healthcare company that has offices distributed nationally, however, there is a very high employee turnover rate. In our results, we found that 55.62% of the end-users has been employed by the company for five years or less. There may be future research done on a more mature, more stable employee base to see if there may be some correlation to the higher vulnerability rates. In addition, being a privately held company, there has historically been less investment in IT security processes, tools, procedures, and training. There may be some differences in outcomes based on company size, stability, and IT security posture. Lastly, as this was a healthcare company, there may be more end-users learnings from other industries or verticals that operate in non-healthcare mediums, which may have impacted the results.

REFERENCES

Antonucci, A. E., Levy, Y., Dringus, L. P., & Snyder, M. (2022). Experimental study to assess the impact of timers on user susceptibility to phishing attacks. *Journal of Cybersecurity Education, Research and Practice, 2021*(No. 2), Article 6.

Carlton, M., & Levy, Y. (2015). Expert assessment of top platform independent cybersecurity skills for non-IT professionals. *Proceedings of the Institute of Electrical and Electronic Engineers Southeast Conference* (pp. 1-6). https://doi.org/10.1109/SECON.2015.7132932

Cooper, M., Levy, Y., Wang, L., & Dringus, L. (2021). Heads-up! An alert and warning system for phishing emails. *Organizational Cybersecurity Journal: Practice, Process and People, 1*(1), 47-68. https://doi.org/10.1108/OCJ-03-2021-0006

Costantino, G., La Marra, A., Martinelli, F., & Matteucci, I. (2018). CANDY: A social engineering attack to leak information from infotainment system. *Proceedings of the IEEE Vehicular Technology Conference, Porto, Portugal* (pp. 1–5).

Dupuis, M. J., Crossler, R. E., & Endicott-Popovsky, B. (2016). Measuring the human factor in information security and privacy. *Proceedings of the IEEE Hawaii International Conference on System Sciences* (pp. 3676-3685).

Earthweb (2022). *How many phishing emails are sent daily in 2022?* https://earthweb.com/how-many-phishing-emails-are-sent-daily/

Federal Bureau of Investigations (FBI) Internet Crime Complaint Center (IC3) (2019, September 10). *Business email compromise the $26 billion scam.* https://www.ic3.gov/media/2019/190910.aspx

Federal Bureau of Investigations (FBI) Internet Crime Complaint Center (IC3) (2020, March 17). *2020 Internet crime report.* https://www.ic3.gov/Media/PDF/AnnualReport/2020_IC3Report.pdf

Federal Bureau of Investigations Internet (FBI) Crime Complaint Center (IC3) (2021, March 23). *2021 Internet crime report*. https://www.ic3.gov/Media/PDF/AnnualReport/2021_IC3Report.pdf

Ho, A. (2018). Rules of three lines of defense for information security and governance. *ISACA Journal, 18*(4), 1-5.

Kost, R. G., & da Rosa, J. C. (2018). Impact of survey length and compensation on validity, reliability, and sample characteristics for ultrashort-, short-, and long-research participant perception surveys. *Journal of Clinical and Translational Science, 2*(1), 31-37. https://doi.org/10.1017/cts.2018.18

Kotson, M., & Shultz, A. (2015). Characterizing phishing threats with natural language processing. *Proceedings of the IEEE Conference on Communications and Network Security*, 308-316. https://doi.org/10.1109/CNS.2015.7346841

Lastdrager, E. E. (2014). Achieving a consensual definition of phishing based on a systematic review of the literature. *Crime Science, 3*(1), 1-10. https://doi.org/10.1186/s40163-014-0009-y

Ramim, M. M., & Lichvar, B. T. (2014). Elicit expert panel perspective on effective collaboration in system development projects. *Online Journal of Applied Knowledge Management, 2*(1), 122-136.

Salahdine, F., & Kaabouch, N. (2019). Social engineering attacks: A study. *Future Internet, 11*(89). https://doi.org/10.3390/fi11040089

Sekaran, U., & Bougie, R. (2016). *Research methods for business: A skill-building approach* (7th Ed.). John Wiley & Sons Ltd.

Walker, A. M., & Selfe, J. (1996). The Delphi method: A useful tool for the allied health researcher. *British Journal of Therapy and Rehabilitation, 3*(12), 677-681. https://doi.org/10.12968/bjtr.1996.3.12.14731.