

Fool Me Once, Shame on You: A Comparison of Generic and Inoculation-Based Warnings for Phishing Email Identification

Imani Bellamy^a, Chih-Wei Hsu^{b,*}, and Zizhao Chen^b

^a*Coppin State University, Baltimore, Maryland, USA*

^b*Department of Computer Science, The University of Texas at Dallas, Texas, USA*

Abstract

Cybercriminals increasingly exploit psychological vulnerabilities and software design weaknesses to deceive users across trusted digital platforms. This study explores patterns among self-reported stress, impulsivity, online behaviors, and phishing-identification performance. Young adults frequently experience emotionally demanding circumstances, including job searches, financial deadlines, and academic stress, that may increase susceptibility to deceptive communications. The study also compares phishing-identification accuracy and confidence among participants who received no warning, a generic warning, or an inoculation-style warning. Grounded in the Heuristic-Systematic Model (HSM) and Cyber-Routine Activity Theory (Cyber-RAT), the study examines how emotional states and routine online behaviors may relate to decision-making during simulated phishing encounters. Participants completed a pre-survey assessing emotional state and online behavior prior to phishing exposure. They were then randomly assigned to receive either no warning, a generic security tip, or a targeted inoculation message designed to reduce susceptibility to phishing. Participants then completed a simulated phishing task using emotionally salient emails (e.g., fake job offers or tuition warnings) to assess trust judgments and response patterns. Participants in the inoculation-warning group showed the highest average accuracy and confidence, while findings involving stress, impulsivity, and online behaviors remained exploratory. By integrating emotional psychology, digital behavior, and preventive intervention strategies, this study provides insight into the psychological factors underlying susceptibility to cyber-enabled fraud. The findings may inform the development of targeted prevention strategies and real-time awareness tools to reduce vulnerability to online deception among young adults.

Keywords: cyber-enabled fraud; phishing susceptibility; psychological vulnerability; heuristic-systematic model (HSM); cyber-routine activity theory (Cyber-RAT); inoculation theory

(Submitted on June 6, 2026; Revised on July 2, 2026; Accepted on August 17, 2026)

© 2026 Totem Publisher, Inc. All rights reserved.

1. Introduction

Phishing is a common form of cyber-enabled fraud. Phishing attacks exploit psychological vulnerabilities through deceptive messages that encourage individuals to disclose sensitive information, click malicious links, or take other actions that compromise their security. Successful phishing attacks can result in financial loss, identity theft, and unauthorized access to personal or organizational accounts. From fake job offers to urgent financial threats, these scams exemplify digital deception designed to exploit psychological vulnerabilities. Young adults may be particularly exposed because frequent online activity, emotional demands, and life transitions can increase both their visibility and vulnerability to attackers. While research has shown that stress and risky online behavior can increase susceptibility [1], few studies have examined whether psychological warning interventions can effectively reduce vulnerability to these attacks.

This study investigates two questions:

- RQ1: What exploratory patterns are observed among self-reported stress, impulsivity, online behaviors, and phishing-identification performance in this sample?
- RQ2: Do participants who receive an inoculation-style warning show higher average phishing-identification accuracy and confidence than those who receive a generic warning or no warning?

* Corresponding author.
E-mail address: cxh210019@utdallas.edu

To examine these questions, the study draws on three psychological frameworks: the Heuristic-Systematic Model (HSM), which explains how emotional pressure leads to snap judgments [1]; Cyber-Routine Activity Theory (Cyber-RAT), which links online habits to risk exposure [1]; and Inoculation Theory, which suggests that exposing people to fake scams in advance can strengthen resistance to real ones [2]. Together, these frameworks provide an integrated lens for examining phishing risk. The findings are important because they help explain not only who may be vulnerable to phishing attacks, but also how targeted warning strategies may reduce susceptibility to online deception. By examining emotional and behavioral patterns associated with phishing susceptibility and comparing specific warning interventions, this study contributes to cybersecurity education and the psychological understanding of online fraud victimization. Ultimately, the study seeks not only to explain why users engage with phishing messages but also to inform interventions that prevent such interactions.

2. Related Work

The rise of cybercrime has underscored the importance of understanding human vulnerability to phishing and spoofing attacks.

The literature reviewed here addresses the psychological, cognitive, and demographic factors that contribute to susceptibility, along with theoretical frameworks and preventive strategies designed to mitigate these risks. This literature review is structured around five central themes: theoretical foundations; phishing and spoofing techniques and psychological tactics; cognitive and demographic vulnerabilities; intervention and prevention; and organizational and large-scale studies.

Theoretical Foundations

Gan, Lee, and Liew apply both Cyber-Routine Activities Theory and the Heuristic-Systematic Model (HSM) to predict phishing susceptibility [1]. Their findings suggest that risky online behaviors significantly increase exposure to phishing, while heuristic processing (i.e., quick, low-effort decision-making) influences how users evaluate suspicious messages. This dual-theory approach highlights the interaction between behavior and cognition, reinforcing the need to address both in phishing interventions. Wu et al. propose inoculation theory as an effective defense against phishing by exposing users to weakened examples of phishing attacks before they encounter real threats [2]. Their findings suggest that preemptive exposure can strengthen users' ability to recognize deceptive tactics and make more informed decisions when evaluating suspicious emails. These theoretical foundations provide the basis for the intervention examined in the present study.

Phishing and Spoofing Techniques and Psychological Tactics

Butavicius et al. and van der Heijden and Allodi examine the psychological manipulation techniques commonly used in phishing emails, including authority, urgency, fear, and other social engineering tactics [3,4]. Their findings demonstrate that even technically knowledgeable users can become victims when attackers exploit emotional responses and cognitive shortcuts. Jayatilaka et al. further show that users' decisions are often influenced by message characteristics, perceived credibility, and decision-making processes rather than carefully evaluating the content of an email [5]. Zhuo et al. synthesize findings from numerous studies in a Systematic Review of Knowledge (SoK), categorizing phishing susceptibility into cognitive, affective, and contextual factors [6]. Together, these studies demonstrate that phishing attacks exploit predictable patterns in human behavior as much as technological vulnerabilities.

Cognitive and Demographic Vulnerabilities

Kavvadias and Kotsilieris provide a wide-ranging review of how demographic variables such as age, education, and gender intersect with psychological traits like impulsivity, trust, and overconfidence [7]. They conclude that young adults and older adults are especially vulnerable, albeit for different cognitive reasons—one due to overconfidence, the other due to cognitive decline.

Frauenstein and Flowerday extend this line of work by using a personality information-processing model to examine how traits such as conscientiousness and neuroticism predict susceptibility [8]. Individuals high in neuroticism are more likely to respond to phishing due to emotional reactivity, while those high in conscientiousness tend to be more cautious. These insights can inform the design of personalized prevention programs.

Intervention and Prevention

Roy and Mukherjee introduce "ShieldUp!", a game-based intervention designed to inoculate users against online scams through interactive learning and repeated practice [9]. Participants who completed the training demonstrated improved scam detection and retained their knowledge longer than those who received traditional awareness materials. Similarly, Wu et al.

demonstrate that inoculation-based phishing interventions can strengthen users' resistance by exposing them to weakened examples of phishing attacks before they encounter real threats [2]. Their findings suggest that theory-driven, preemptive training can improve users' ability to recognize and resist phishing attempts. They further emphasize the importance of continuously evaluating and improving security awareness programs to promote long-term changes in user behavior rather than relying solely on one-time training sessions [10].

Organizational and Large-Scale Studies

Rizzoni et al. evaluated a large-scale phishing simulation conducted within a hospital setting and found that realistic phishing simulations can improve employee awareness while identifying areas where additional training is needed [11]. Cross and Holt examined the broader consequences of online fraud victimization, demonstrating that cybercrime can extend beyond financial loss to include identity theft and other long-term harms [12]. Together, these studies emphasize that effective phishing prevention requires both realistic organizational training and a broader understanding of the long-term impact of cyber-enabled fraud.

This review highlights a multidimensional framework for understanding susceptibility to phishing and spoofing: one that incorporates cognitive psychology, behavioral science, and social engineering tactics. The literature suggests that effective intervention requires a layered defense that is grounded in theory, sensitive to user psychology, and integrated into organizational ecosystems. These findings inform the design and rationale of the current study, which explores a psychologically grounded, user-informed approach to reducing phishing susceptibility.

3. Methods

A total of 85 responses were collected through the online survey. After removing duplicate or invalid responses, the final analytic sample was $N = 78$. The sample included Research Experience for Undergraduates (REU) interns and other respondents. Participation was voluntary. Participants ranged in age from 16 to 59 and represented diverse gender identities and college-enrollment statuses. Among the 78 respondents, 45 identified as male, 31 as female, and 2 as non-binary. Most participants were currently enrolled in college, while a smaller portion reported not being in college.

Duplicate submissions and responses associated with reused participant IDs were identified during data cleaning and excluded before analysis.

Design

The research employed a between-subjects experimental design with three groups:

- Group A (Control Group) – Received no warning before evaluating the emails.
- Group B (Subtle Warning Group) – Received a brief, general reminder to remain cautious about phishing, without examples of specific phishing tactics or guidance on how to resist them.
- Group C (Inoculation Warning Group) – Received an inoculation-based warning that presented an example of a common phishing tactic together with a counterargument explaining how to recognize and resist that tactic.

Participants were randomly assigned to one of the three groups ($n = 25$ in Group A, $n = 26$ in Group B, and $n = 27$ in Group C). Each participant was given a unique ID (e.g., P001, P002) to protect confidentiality and organize responses across spreadsheets.

Materials

All participants viewed the same set of 10 emails, comprising a mix of phishing and legitimate messages. The emails were presented as screenshots in a consistent visual format. Email topics included fake scholarship offers, package delivery notices, and job opportunities. The emails were edited to maintain a uniform appearance while preserving realistic phishing elements such as suspicious links, generic greetings, and urgent language.

Before the phishing task, all participants completed a detailed pre-survey assessing the following:

- Demographics: Age, gender, college status.
- Emotional and situational status: Self-reported calm/stress levels and recent stress indicators (e.g., job/financial concerns, scholarship anxieties).
- Digital behavior: Frequency of email checking, susceptibility to clicking suspicious links, multitasking habits, and previous exposure to phishing.
- Psychological traits: Impulsivity, emotional reactivity, tendencies toward worry, and confidence in identifying phishing emails.
- Self-efficacy: Items gauging participants' self-perceived phishing detection ability and skepticism levels.

The Google Forms survey included Likert-scale, multiple-choice, and open-response items.

Participants responded to the following three questions per email: (1) Would you open this email? (Yes / No / Unsure), (2) Do you believe this email is legitimate or trustworthy? (legitimate / Not legitimate / Unsure), and (3) How confident are you in your decision? (Scale from 1–10). Responses were collected via Google Forms and automatically sorted into Google Sheets, with a separate tab for each experimental group. This approach enabled simplified tracking of individual responses and group averages.

Procedure

The study was originally designed for in-person administration in a supervised setting, such as a classroom, to standardize completion conditions. Because of scheduling and logistical constraints, it was instead distributed through a shareable Google Forms link. Participants completed the survey on their own time and using their own devices, which introduced variability in response conditions. Participants were instructed to complete the pre-survey before opening the simulated phishing task. The pre-survey collected demographic information, emotional and situational measures, digital-behavior items, and psychological-trait items. After submitting the pre-survey, participants proceeded to the email-evaluation form containing the 10 email screenshots.

Participants were randomly assigned to one of the three experimental conditions and were not informed of their group labels. The assigned warning (or no warning) appeared at the start of the email-evaluation form, while the researcher tracked condition assignments. All Google Forms items were set as required, reducing item-level missingness among submitted responses. This setting prevented item-level omissions within completed submissions, although it did not prevent duplicate records or reused IDs; those issues were addressed separately during data cleaning. Each participant entered a unique ID, and completed responses were transferred to group-specific Google Sheets for analysis.

4. Results

To evaluate participant performance, accuracy and confidence scores were calculated through a structured multi-step process. During the simulated phishing task, participants evaluated 10 distinct emails and indicated whether each message was real (legitimate), fake (phishing), or if they were unsure. One accuracy point was awarded when a participant correctly identified an email; an incorrect or "unsure" response received 0 points. Each participant's correct responses were summed to produce an individual accuracy score out of 10. For each condition, mean accuracy was calculated by summing the individual accuracy scores and dividing by the number of participants in that condition (e.g., Group A, $M = 6.68$). For confidence, participants rated their certainty for each email decision on a 1-to-10 Likert scale. Each participant's mean confidence score was calculated by summing the 10 item-level confidence ratings and dividing by 10. Group-level mean confidence was then calculated from the individual mean confidence scores within each experimental group. (See Figure 1 and Table 1 below for average accuracy scores by group; Figure 2 and Table 2 for average confidence scores.)

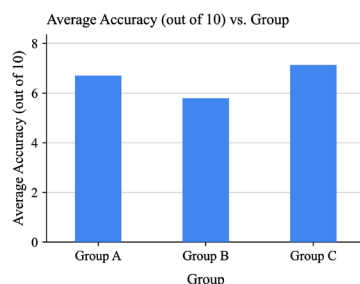


Figure 1. Bar graph shows the average accuracy scores (out of 10) for each group in the phishing detection study (Group C had the highest accuracy, followed by Group A, with Group B scoring the lowest)

Table 1. Numerical data showing the average accuracy (out of 10) for Groups A, B, and C

Group (letter)	Average Accuracy (out of 10)
Group A	6.68
Group B	5.77
Group C	7.11

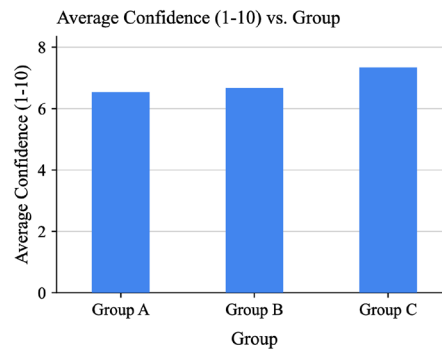


Figure 2. Bar graph showing the average confidence scores (out of 10) for each group in the phishing detection study (Group C had the highest confidence, followed by Group B and then Group A)

Table 2. Numerical data showing the average confidence (out of 10) for Groups A, B, and C

Group	Average Confidence (out of 10)
Group A	6.51
Group B	6.65
Group C	7.32

Visualization

Figures 1 and 2 summarize group-level trends in accuracy and confidence, respectively. Figure 1 displays average accuracy by group. Group C, which received the inoculation message, had the highest accuracy (7.11), while Group B, which received a subtle warning, had the lowest (5.77). Figure 2 presents average confidence by group. Group C again had the highest average confidence (7.32), followed by Group B (6.65) and Group A (6.51). Although Group C achieved the highest average phishing detection accuracy ($M = 7.11$), it also showed the smallest standard deviation ($SD = 1.48$), indicating that participants in this group performed more consistently than those in the other two groups. Group A had an average accuracy of 6.68 ($SD = 2.32$), while Group B had the lowest average accuracy of 5.77 ($SD = 2.16$), suggesting greater variability in participant performance. Confidence scores followed a similar pattern. Group C reported the highest average confidence ($M = 7.32$, $SD = 1.25$), followed by Group B ($M = 6.65$, $SD = 1.33$) and Group A ($M = 6.51$, $SD = 1.15$). Descriptively, Group C achieved the highest mean accuracy and the least variability in accuracy. Because this was a pilot study, group differences are interpreted descriptively rather than as statistically significant effects. Although Group C demonstrated higher average accuracy and confidence scores than the other groups, no inferential statistical analyses were conducted to determine whether these differences were statistically significant. Therefore, these findings should be interpreted as descriptive trends rather than conclusive evidence of group differences.

Patterns in Individual Data

Participant-level data within each group revealed several patterns:

- In Group C, higher average accuracy was accompanied by higher average confidence at the group level. One possible interpretation is that the inoculation message provided more specific guidance for evaluating phishing cues, although the study did not test this mechanism directly.
- In Group B, several participants reported relatively high confidence despite lower accuracy scores, indicating descriptive gaps between perceived and actual performance in individual cases. The generic warning may have increased vigilance without providing specific decision criteria; however, the study did not directly examine participants' reasoning or test whether the warning produced a false sense of security.

In Group A, confidence levels were generally lower and more closely aligned with accuracy, indicating smaller descriptive confidence–accuracy gaps than those observed in Group B. This pattern may reflect greater uncertainty in the absence of warning guidance, although the study did not directly measure participants' reliance on intuition or self-doubt.

Taken together, these descriptive patterns suggest closer alignment between confidence and accuracy in Group C, while several participants in Group B showed larger confidence–accuracy gaps.

Analysis of Impulsivity, Digital Behavior, and Emotional Targeting

Impulsivity and emotional reactivity were examined as descriptive baseline characteristics because rapid, cue-driven responding may be relevant when users evaluate urgent or emotionally salient messages. The composite was used to compare baseline patterns across the three experimental groups; it was not treated as a validated clinical or diagnostic measure.

Likert-type items were used because they allow respondents to indicate the degree to which each statement reflects their typical reactions rather than forcing a binary response. In this study, the five items provided a structured way to summarize self-reported impulsive tendencies and emotional reactivity across participants. The resulting composite was intended for descriptive comparison within this sample, not for clinical assessment or diagnosis.

To examine these measured characteristics, an impulsivity-and-emotional-reactivity composite was calculated from five pre-survey Likert-type items. Average composite scores were then calculated for each experimental group. As shown in Figure 3, the average scores were similar across groups: Group A ($M = 14.36$, $SD = 2.84$), Group B ($M = 13.38$, $SD = 2.87$), and Group C ($M = 14.78$, $SD = 3.12$). These descriptive averages do not show a strong group-level difference in impulsivity. Because no inferential or participant-level association test was conducted, this study does not determine whether impulsivity contributed to the observed differences in phishing-detection performance.

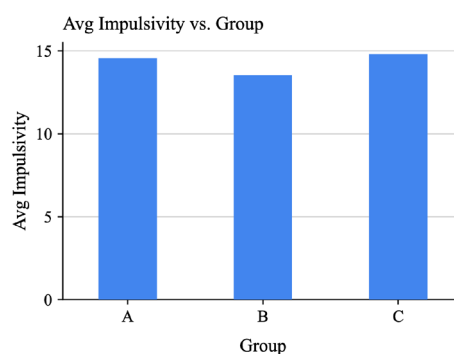


Figure 3. Bar graph displaying average impulsivity scores across experimental groups (A, B, and C). The graph illustrates minimal variation in impulsivity levels between groups.

To further address the first research question, the pre-survey data were examined for patterns involving stress, impulsivity, and online behavior. The pre-survey included five questions related to impulsivity and emotional reactivity, such as:

- “I often act without thinking.”
- “I tend to worry about small things.”
- “I’m good at controlling my impulses.”
- “I feel emotionally reactive to messages or notifications.”
- “I often rush through tasks or messages to get them done.”

Each item was rated from Strongly Disagree (1) to Strongly Agree (5). Four items were scored so that higher values indicated greater impulsive tendencies or emotional reactivity. The self-control item (“I’m good at controlling my impulses”) was reverse-scored so that all items pointed in the same direction. This step ensured that agreement with the positively worded self-control item was not incorrectly interpreted as greater impulsivity when the five responses were summed. The theoretical composite range was 5–25; observed scores in this sample ranged from 8 to 23. Higher scores indicated greater impulsive tendencies and emotional reactivity.

The impulsivity scores were then examined alongside each participant’s average phishing accuracy and confidence scores. The similarity of the group-level impulsivity averages suggests that the observed performance pattern was unlikely to be driven by substantial baseline differences in this measured characteristic. Against this broadly comparable baseline, Group C’s higher average accuracy and confidence are consistent with the intended benefit of the inoculation-based warning. Inoculation theory suggests that exposing individuals to weakened forms of manipulation or providing preemptive warnings can help build resistance to future deceptive attempts. The pre-survey also included exploratory questions related to stress and

online behavior. While impulsivity was examined quantitatively through a composite score, the stress and online behavior items were used to identify behavioral patterns that may contribute to phishing susceptibility rather than to test formal statistical relationships.

The questions related to digital behavior were:

- “How often do you check your email per day?”
- “How often do you click on links without verifying the source?”
- “Do you multitask while reading messages?”

Descriptive analysis revealed a consistent pattern: participants who reported checking email more frequently, clicking links quickly, or multitasking tended to show lower phishing-detection performance. Although these relationships were not evaluated using inferential statistics, the convergence of the observed patterns suggests that routine online behaviors may serve as meaningful indicators of phishing risk and warrants confirmatory testing in a larger sample.

To reflect these concerns in the experiment, the simulated phishing task used email topics representing real-world emotional triggers for young adults. Everyone saw the same 10 emails, which included:

- Fake job offers
- Scholarship announcements
- FAFSA warnings
- Package delivery notices (used as filler)

These topics were chosen because they align with the emotional timing and public concerns that many young adults face, especially during periods of school stress or financial deadlines. This design simulated phishing messages that use emotional urgency and predictable user routines, such as job hunting or checking email between classes. Together, the pre-survey and phishing simulation suggest that emotional impulsivity and routine online behaviors may be associated with phishing susceptibility. Although these exploratory findings indicate potential behavioral patterns, no formal statistical analyses were conducted to establish causal relationships. These findings support the idea that phishing attacks may exploit predictable emotional situations and routine online behaviors, but additional research is needed to confirm this.

5. Discussion

Group C achieved higher average phishing-detection accuracy and confidence than Groups A and B, providing preliminary support for the potential value of inoculation-based warnings as a targeted phishing-prevention strategy. These findings align with prior studies on Inoculation Theory, which argue that exposing individuals to weakened persuasive messages and counterarguments can build resistance to future manipulations. In contrast, Group B had the lowest average accuracy and reported mid-level confidence. This descriptive pattern suggests that generic warnings may provide less actionable guidance than targeted inoculation messages, although additional testing is needed to determine whether this difference represents a reliable warning effect. Group A, which received no warning, performed moderately well, suggesting that some participants may rely on intuition or prior experience in the absence of priming. This pattern suggests that intuition or prior experience may support some phishing judgments, whereas targeted guidance may provide more explicit criteria for evaluating suspicious messages. Future studies could vary the difficulty and subtlety of phishing cues to test whether targeted guidance is especially useful when deceptive features are less obvious.

Average impulsivity scores were similar across the three groups, suggesting that substantial group-level differences in measured impulsivity were unlikely to account for the observed performance pattern. Group C nevertheless achieved the highest average accuracy and confidence, reinforcing the potential value of the inoculation-based warning. Research supports that impulsivity can increase risky cybersecurity behaviors, such as clicking on suspicious links, but it does not always fully predict who will fall for phishing attacks. The phishing scenarios were designed around emotionally salient situations for young adults, including job opportunities, financial-aid concerns, and academic deadlines. Although emotional context and timing were not independently manipulated, the observed response patterns are consistent with the possibility that emotionally relevant cues may influence phishing judgments. Regarding the second research question, Group C achieved the highest average phishing-detection accuracy ($M = 7.11$) and confidence ($M = 7.32$), despite similar average impulsivity scores across the three groups. These descriptive findings provide preliminary support for the potential effectiveness of inoculation-based warnings in this study, although inferential testing is needed to confirm the reliability of the observed group differences.

Regarding RQ1, the study identified exploratory patterns involving impulsivity and routine online behaviors, but it did not establish that stress, impulsivity, or online behaviors increased phishing susceptibility. Group-level impulsivity scores were similar across the three conditions, while frequent email checking, rapid link clicking, and multitasking were descriptively associated with lower phishing-identification performance. Regarding RQ2, participants in the inoculation-warning condition had the highest average phishing-identification accuracy and confidence. These descriptive results provide preliminary support for inoculation-based warnings, although inferential testing is needed to determine whether the warning reliably caused the observed differences.

6. Limitations and Future Work

This study had several limitations. First, the relatively small sample size of 78 participants limited the statistical power of the study, so the findings should be interpreted as suggestive rather than definitive. In addition, because the survey was completed remotely, participants may have encountered distractions or varied in the level of attention they devoted to the task. The sample also had limited demographic diversity. Although participants varied in age and gender, most were college students or interns, which may restrict the generalizability of the findings to broader populations. Furthermore, the five-item composite measure combined self-reported impulsive tendencies, emotional reactivity, worry, and self-control. As a result, it should be viewed as a broad descriptive indicator rather than a validated measure of a single psychological construct. The study also used only 10 email stimuli drawn from a limited range of topics, meaning that the results may have been influenced by the specific balance of legitimate and phishing messages or by the salience of particular cues. Finally, the analysis relied primarily on descriptive comparisons of group means and did not examine participant-level relationships among impulsivity, confidence, and accuracy. Larger samples and inferential analyses would therefore be needed to estimate uncertainty more precisely and determine whether observed differences reflect intervention effects rather than random group variation.

Future research could address these limitations in several ways. Conducting the study in a classroom or controlled laboratory setting could provide more consistent testing conditions, improve the accuracy of timing, and reduce the likelihood of duplicate responses. Brief follow-up surveys or interviews could also provide deeper insight into participants' decision-making processes and how they interpret phishing-related cues. In addition, future studies should use an established and validated psychological scale and report its reliability within the study sample. Expanding the stimulus set to include a larger, pretested collection of emails covering financial, academic, social, and other common phishing contexts would also create a more realistic testing environment and strengthen the generalizability of the findings. Finally, longitudinal testing could examine whether the effects of inoculation warnings persist over time, such as after one week or one month, providing a clearer understanding of their durability and potential value for long-term cybersecurity training.

7. Conclusion

This study met its core objective by evaluating whether inoculation-style warnings could improve phishing-detection accuracy and user confidence. Despite logistical challenges, the experimental design, survey structure, and analysis yielded meaningful insights. In particular, Group C's higher average accuracy and confidence highlight the potential value of targeted warning messages grounded in psychological theory. Although the pre-survey findings were primarily descriptive, they provided useful context for understanding participant attitudes and behavioral patterns. Many participants reported general awareness of phishing but limited detailed knowledge, underscoring the continuing need for educational interventions. Future work should formally analyze the pre-survey data to examine associations among emotional state, impulsivity, and phishing susceptibility and to identify potential predictors that may inform targeted warning interventions. Such analyses could inform future targeted warning interventions. Moreover, this project demonstrated the feasibility of adapting complex psychological models (like Inoculation Theory and the Heuristic-Systematic Model) into short-form digital interventions. Even a brief paragraph presenting a persuasive technique and counterargument was associated with higher average phishing-detection accuracy and confidence in this study. The study also identified practical considerations for future data collection, survey administration, and data organization. Group-specific data organization and automated calculation of average scores facilitated comparisons across conditions and can support future iterations with larger, more controlled samples.

Acknowledgments

This work was supported in part by the U.S. National Science Foundation under Grant 2349347. It was partially completed by Imani Bellamy under the supervision of Professor W. Eric Wong and Mr. Chih-Wei Hsu while attending the Research Experiences for Undergraduates (REU) program at the University of Texas at Dallas in the Summer of 2025.

References

- [1] Gan, C.L., Lee, Y.Y., and Liew, T.W., 2024. Fishing for phishy messages: predicting phishing susceptibility through the lens of cyber-routine activities theory and heuristic-systematic model. *Humanities and Social Sciences Communications*, 11(1), 1552.
- [2] Wu, D., Zhang, J., Brown, N., Lowry, P.B., and Moody, G.D., 2020. Patching the “human” in information security: using the inoculation defense to confer resistance against phishing attacks.
- [3] Butavicius, M., Parsons, K., Pattinson, M., and McCormac, A., 2016. Breaching the human firewall: social engineering in phishing and spear-phishing emails. *Arxiv Preprint Arxiv:1606.00887*.
- [4] Van Der Heijden, A., and Allodi, L., 2019. Cognitive triaging of phishing attacks. In *28th USENIX Security Symposium (USENIX Security 19)*, pp. 1309-1326.
- [5] Jayatilaka, A., Arachchilage, N.A.G., and Babar, M.A., 2024. Why people still fall for phishing emails: an empirical investigation into how users make email response decisions. *Arxiv Preprint Arxiv:2401.13199*.
- [6] Zhuo, S., Biddle, R., Koh, Y.S., Lottridge, D., and Russello, G., 2023. SoK: human-centered phishing susceptibility. *ACM Transactions on Privacy and Security*, 26(3), pp. 1-27.
- [7] Kavvadias, A., and Kotsilieris, T., 2025. Understanding the role of demographic and psychological factors in users’ susceptibility to phishing emails: A review. *Applied Sciences*, 15(4), 2236.
- [8] Frauenstein, E.D., and Flowerday, S., 2020. Susceptibility to phishing on social network sites: A personality information processing model. *Computers & Security*, 94, 101862.
- [9] Roy, A., and Mukherjee, S., 2025. Shieldup!: inoculating users against online scams using a game based intervention. *Arxiv Preprint Arxiv:2503.12341*.
- [10] Jayatilaka, A., Beu, N., Baetu, I., Zahedi, M., Babar, M.A., Hartley, L., and Lewinsmith, W., 2021. Evaluation of security training and awareness programs: review of current practices and guideline. *Arxiv Preprint Arxiv:2112.06356*.
- [11] Rizzoni, F., Magalini, S., Casaroli, A., Mari, P., Dixon, M., and Coventry, L., 2022. Phishing simulation exercise in a large hospital: A case study. *Digital Health*, 8, 20552076221081716.
- [12] Cross, C., and Holt, T.J., 2023. More than money: examining the potential exposure of romance fraud victims to identity crime. *Global Crime*, 24(2), pp. 107-121.