Asian Journal of Science and Applied Technology
ISSN: 2249-0698 (Print)
Journal Homepage: https://ajsat.org/
DOI: https://doi.org/10.70112/ajsat-2026.15.1.4350

Research Article

Measuring the Impact of Embedded Phishing Training on Ransomware-Lure Click Behavior: A Field Study Across Three Organizations

Aminu Muhammad Auwal[1*] and Sulaiman Shehu[2]
[1*]Faculty of Natural Sciences, University of Jos, Plateau State, Nigeria
[2]Department of Software Engineering, Bayero University, Kano, Kano State, Nigeria

Article Information

Article History
Received: 10 March 2026
Revised: 10 April 2026
Accepted: 25 April 2026
Published online: 5 May 2026

Keywords
Phishing Simulation
Embedded Training
Ransomware Delivery
Field Experiment
Randomized Controlled Trial
Click-Through Rate
Lure Category
Urgency
Engagement
Time-On-Module
Protection Motivation Theory
Security Awareness

Correspondence*
i.elameenu@gmail.com

ORCID
Aminu Muhammad Auwal
https://orcid.org/0009-0005-1799-7876
Sulaiman Shehu
https://orcid.org/0009-0002-0056-604X

Abstract

Embedded anti-phishing training, redirecting employees who fail simulated phishing campaigns to immediate instructional content, is among the most widely deployed behavioral defenses against ransomware-delivering social engineering attacks. Despite this widespread adoption, recent large-scale randomized evaluations have produced evidence of only marginal absolute reductions in failure rates, raising fundamental questions about whether the current model of simulated phishing plus embedded training is an effective and efficient use of organizational security resources. A critical gap in the existing literature is the absence of lure-category-specific analyses: virtually all prior evaluations have used heterogeneous phishing campaign portfolios without examining whether training effectiveness varies systematically across the specific delivery techniques—fake software updates, spoofed invoices, and urgency-framed credential resets—most commonly associated with ransomware campaigns. This study aimed to: (1) determine whether embedded 90-second micro-training produces a statistically significant reduction in click-through rates for ransomware-style phishing lures compared with instructor-led training only and a no-training control in a randomized field experiment; (2) examine whether training effectiveness differs across three ransomware-relevant lure categories differing in their reliance on visual artifact inspection versus urgency-driven cognitive override; and (3) assess whether employee engagement with embedded training content as measured by time-on-module moderates the protective effect at the individual level. A six-month randomized field experiment was conducted across three mid-sized organizations (combined n = 1,284 employees). Employees were randomly assigned within department strata to embedded micro-training (n = 428), instructor-led annual training only (n = 426), or a control group receiving standard onboarding only (n = 430). Six monthly simulated campaign waves sent one email from each of three ransomware-relevant lure categories—fake software update notifications, spoofed invoices, and urgent credential-reset requests—to all participants. Click-through rates were compared within conditions using McNemar's test and between conditions using chi-square tests. Individual-level engagement (time-on-module) was assessed as a moderator among the embedded training group. The embedded micro-training group showed a significant overall click-rate reduction from 18.4% to 13.1% (−5.3 pp, p = 0.002), significantly larger than the instructor-led (−1.5 pp, p = 0.214) or control (−1.7 pp, p = 0.389) reductions. Effectiveness was highly lure-dependent: the fake software update category showed a 12.4 pp reduction (p < 0.001); the spoofed invoice category showed a 4.4 pp reduction (p = 0.038); the urgent credential-reset category showed only a 1.8 pp non-significant reduction (p = 0.412). Participants engaging with training content above the median time (>52 seconds) showed significantly larger individual-level reductions in subsequent click likelihood (OR = 0.52, p = 0.011) than below-median engagers (OR = 0.81, p = 0.204). Embedded micro-training produces a measurable but strongly lure-dependent reduction in ransomware-relevant phishing susceptibility. Urgency-based credential-reset lures, among the most common ransomware initial-access vectors, are largely resistant to this intervention, indicating that technical controls including multi-factor authentication and DNS-level phishing filtering are necessary complements regardless of training investment. Engagement with training content rather than mere completion is a key moderator of effectiveness, suggesting that addressing the engagement gap may be as important as content quality in improving training outcomes.

I. INTRODUCTION

Phishing and social engineering attacks are the dominant initial access vectors for ransomware infections, accounting for a large majority of confirmed incidents across sectors and geographies [1], [2]. Organizations have responded to this threat primarily by deploying simulated phishing campaigns, in which employees receive realistic but harmless test emails and are tracked for click-through behavior, combined with embedded training, in which employees who fail a simulation are immediately redirected to short instructional content explaining the warning signs they missed [3].

This model of test-and-train has become the standard offering of commercial security awareness training platforms and is required by cybersecurity compliance frameworks in a growing number of regulated sectors. The theoretical appeal of embedded training is grounded in learning science principles: immediate feedback provided at the moment of error, in a context directly relevant to the behavior being trained, should produce stronger learning effects than decontextualized classroom instruction delivered at a temporal remove from the behavior of interest [4]. Protection Motivation Theory (PMT) provides an additional theoretical basis: effective training should increase both threat appraisal (recognition that the email represents a real risk) and coping appraisal (confidence in the ability to identify and appropriately respond to the threat), thereby increasing the likelihood of protective behavior when the next lure is encountered [5].

Despite this strong theoretical rationale, empirical evidence for the effectiveness of the embedded training model has been consistently weaker than expected. The most rigorous published evaluation, an eight-month randomized controlled trial across more than 19,500 employees at a large healthcare organization, found that embedded training produced only a 1.7 percentage point absolute reduction in phishing failure rates, and that prior annual security awareness training completion showed no significant relationship with simulation performance [6], [7].

A field study of 191 employees found that urgency-framed lures significantly increased susceptibility regardless of whether employees had received training [8]. A scoping review of 42 studies found that near-term training effects are better documented than long-term behavioral change, and identified design factors including training intensity, active learning, and detailed feedback as the key determinants of outcome, but noted that a substantial minority of studies found null or even negative training effects [9].

Two important gaps remain in this literature. First, virtually all prior studies have evaluated training effectiveness using heterogeneous phishing campaigns that mix lure types without examining whether training effects vary across specific lure categories. Ransomware delivery campaigns rely on a relatively small set of documented lure archetypes—fake software updates, spoofed invoices, and urgency-based credential requests—which differ substantially in their cognitive processing demands and the degree to which explicit knowledge of warning signs can interrupt automatic processing [10], [11].

Second, the role of training engagement—how much time and attention employees devote to embedded training content when it is presented—has not been systematically examined as a moderator of training effectiveness, despite a well-established literature on engagement as a predictor of learning outcomes across educational contexts. The present paper addresses both gaps through a six-month, three-arm randomized field experiment across three mid-sized organizations, using simulated phishing campaigns modeled specifically on documented Locky, Ryuk, and Conti-style ransomware delivery lures. The experiment is designed to: (1) test whether embedded micro-training produces a significantly larger reduction in click-through rates than instructor-led training only or no training; (2) compare training effectiveness across three ransomware-relevant lure categories differing in their reliance on visually inspectable artifact cues versus urgency-driven cognitive override; and (3) examine time-on-module as a moderator of individual-level training effectiveness.

II. RELATED WORK

A. Theoretical Framework: PMT and Dual-Process Theory

The theoretical framework for the present study integrates Protection Motivation Theory (PMT) with dual-process accounts of phishing susceptibility. PMT proposes that protective behavior is motivated by two interacting cognitive appraisal processes: threat appraisal, which evaluates the perceived severity and probability of a threat, and coping appraisal, which evaluates the individual's belief in their ability to successfully implement the recommended protective response [5]. Applied to phishing susceptibility, threat appraisal corresponds to the recognition that an incoming email may be a phishing attempt, that the threat is real and personally relevant, while coping appraisal corresponds to confidence in the ability to correctly identify the specific warning signs present in the email and take appropriate action [12].

Dual-process theories of cognition distinguish between deliberate, effortful System 2 processing and fast, automatic, heuristic-based System 1 processing. Phishing lures, including ransomware delivery emails, are designed to exploit System 1 processing by creating conditions—urgency, authority, and familiarity—under which recipients respond automatically rather than deliberately evaluating the email's authenticity [10], [13].

Embedded training, by highlighting the specific visual and contextual warning signs present in a lure at the moment of failure, aims to strengthen the cue-recognition schema that triggers System 2 processing when a potentially suspicious email is encountered. The prediction following from this

framework is that training should be most effective for lure categories that contain visually inspectable artifacts—mismatched sender domains, unusual attachment types, and branding inconsistencies—and least effective for urgency-exploiting lures that produce System 1 responses powerful enough to override trained cue-recognition schemas, a prediction directly tested in the present experiment.

B. Prior Randomized and Field Evaluations

The landmark randomized evaluation by Ho *et al.* [6], [7] remains the most methodologically rigorous published assessment of embedded phishing training effectiveness. Conducted at a large US healthcare organization with over 19,500 participants across an eight-month period, the study found that recipients of embedded training showed a 1.7 percentage point lower failure rate than untrained recipients in subsequent campaigns, a result that, while statistically significant in the large sample, is of questionable practical significance. Notably, the study found no significant association between completion of annual instructor-led security awareness training and phishing simulation performance, suggesting that the instructor-led model alone is essentially ineffective.

A field study with 191 employees at an Italian organization examined the comparative effectiveness of authority-based versus urgency-based phishing lures, finding that urgency-framed lures produced significantly higher click rates regardless of employee demographic characteristics, and that embedded training, while perceived positively by employees, did not measurably reduce failure rates [8]. A repeated-measures experiment found that the specific scenario presented in a phishing simulation was the strongest predictor of whether an individual would click, substantially outweighing the effects of prior training, influence technique, and simulated sender personalization, and that individual-level susceptibility was substantially unpredictable across different lure types presented to the same employee [11].

A scoping review of 42 phishing training studies found that active learning approaches, high training intensity, and detailed contextualized feedback were the design elements most consistently associated with positive training outcomes [9]. The review also found that succinct in-line warning mechanisms on suspect emails and easy-to-use reporting mechanisms with positive feedback were effective complements to training content [9].

Critically, the review noted that some studies reported training paradoxically increasing susceptibility—attributed to overconfidence effects in which trained employees become more certain of their ability to identify phishing and therefore apply less deliberate scrutiny [9]. The ethical responsibilities of security professionals designing and deploying these training systems extend to avoiding such counterproductive outcomes through careful evaluation and evidence-based program design [14].

C. Ransomware Delivery Lure Categories

Documented ransomware delivery campaigns have relied on a small number of consistently effective lure categories. Fake software update and license renewal notifications, mimicking communications from widely used collaboration, productivity, or security software vendors, exploit users' normalized expectations of receiving such notifications and their association with a trusted vendor brand [1], [2]. These lures typically contain visually inspectable artifacts including sender domain mismatches, branding inconsistencies, and unexpected attachment file types that should, in principle, be identifiable by trained users. Spoofed invoice and payment-request emails exploit organizational financial workflows and the normalized expectation of receiving and processing invoices, particularly in accounting and procurement roles [1].

Urgency-based credential-reset and account-compromise notifications exploit a different cognitive mechanism. Rather than relying on visual artifact inspection, these lures create a negative emotional state—fear of account loss and anxiety about a security breach—that activates threat appraisal while simultaneously undermining coping appraisal by implying that immediate action is required before the user has time for deliberate evaluation [10], [12]. This manipulation of the PMT process explains why urgency-based lures are disproportionately effective: they simultaneously raise threat salience (increasing motivation to respond) and reduce the time available for coping appraisal (reducing the likelihood that the response involves proper authentication verification). Phishing detection technologies, including DNS and IP filtering [15] and web ontology-based user profiling [16], provide technical defenses that operate on the delivery infrastructure rather than on the cognitive appraisal process, and therefore are not subject to the urgency-override limitation that constrains behavioral training.

D. Engagement as a Moderator of Training Effectiveness

A well-established literature in educational psychology and learning science documents that time-on-task and active engagement with learning content are among the strongest predictors of knowledge acquisition and behavioral transfer [4]. In the context of embedded security awareness training, a key practical concern is that employees who are redirected to training content after failing a simulation may treat the training module as an administrative obstacle to be bypassed as quickly as possible, clicking through the minimum number of required interactions without engaging with the content. If a substantial proportion of triggered training instances produce minimal engagement, the aggregate training effect observed at the population level may understate the potential effect achievable with higher engagement, and strategies for increasing engagement may offer more leverage than further refinement of training content. The present study measures time-on-module as a proxy for engagement and examines its moderating effect on individual-level subsequent susceptibility.

Criminological perspectives on victimization suggest that employee engagement with security training may also be moderated by the perceived relevance of training content to the employee's own risk context, with employees who have personally experienced or witnessed attempted social engineering attacks showing higher engagement and better transfer than those who perceive the threat as abstract [17]. The ethical obligation of security professionals to design training that is genuinely protective rather than merely compliance-satisfying includes attention to the engagement dimension of training effectiveness [14].

III. METHODOLOGY

A. Participating Organizations and Ethical Approval

Three mid-sized organizations agreed to participate after being approached through a professional cybersecurity association: a regional healthcare network (412 employees), a manufacturing firm (487 employees), and a professional services firm (385 employees), totaling 1,284 employees across all three sites. All organizations had existing commercial phishing-simulation platforms in use for compliance purposes and agreed to allow the research team to control campaign content and randomization for the study period. Ethical approval was obtained from the University's Research Ethics Committee. Participating employees were notified at the outset of the study period that their organization would be conducting enhanced phishing simulation activities as part of a research collaboration, consistent with standard practice for simulated phishing research, but were not informed of their individual condition assignment. Data were collected and analyzed in anonymized aggregate form.

B. Randomization and Condition Assignment

Within each organization, the full employee roster was stratified by department and seniority level, and employees were randomly assigned within strata to one of three conditions using a computer-generated allocation sequence: (1) embedded micro-training (EMT, n = 428 across all organizations); (2) instructor-led annual training only (ILT, n = 426); or (3) control, receiving standard onboarding security content only (CTL, n = 430). Stratified randomization was used to balance role-based differences in email volume, organizational authority level, and prior exposure to phishing simulations across conditions. Allocation was concealed from line managers and from participants; only the research team and IT administrators who configured the simulation platform had access to condition assignments.

C. Campaign Design and Lure Construction

Six monthly simulated phishing campaign waves were conducted over the six-month study period (Waves 1–6). Each wave sent one email from each of three lure categories to all 1,284 participants, for a total of three emails per participant per wave and eighteen emails per participant over the study period. The three lure categories were (i) Fake Software Update (FSU): notifications mimicking update or license renewal alerts from common collaboration, productivity, and security software vendors, including a sender domain that differed by one character from the legitimate domain, an urgency prompt, and a link or attachment with an unexpected file extension; (ii) Spoofed Invoice (SI): payment-request or vendor invoice emails with a malicious attachment placeholder, sender domain mismatch, and slightly unfamiliar but plausible vendor name; and (iii) Urgent Credential Reset (UCR): emails claiming that the recipient's account had been compromised or that a mandatory policy-driven password reset was required within two hours, featuring a cloned corporate identity, high urgency language, and a link to a credential-harvesting page replica. Lure templates were constructed by the research team from documented characteristics of confirmed ransomware delivery campaigns attributed to LockBit, BlackCat/ALPHV, and Cl0p affiliates. Sending time, sender name, and subject line variants were rotated across waves to reduce template-specific familiarity effects.

D. Training Conditions

Participants in the EMT condition who clicked a simulated phishing link were immediately redirected to a 90-second interactive training module specific to the lure category that triggered the redirect. Each module was designed around the specific visual or linguistic warning signs present in the lure. For the FSU lure, the module highlighted the sender domain mismatch, the unexpected attachment extension, and the urgency framing. For the SI lure, it highlighted the unfamiliar vendor name and the attachment request outside normal invoice workflow. For the UCR lure, it highlighted the urgency language and the link destination domain mismatch. Each module included two comprehension check interactions that required the user to identify the warning signs before proceeding. Module completion and time-on-module were logged by the platform for each triggered instance. EMT participants who did not click a given lure received no additional content for that wave.

Participants in the ILT condition received a single 45-minute instructor-led training session at the start of the study period (before Wave 1). The session covered general phishing awareness across all three lure categories, the organization's reporting procedures, and common persuasion techniques used in social engineering attacks. No further formal training was provided to ILT participants during the six-month campaign period. CTL participants received no additional training beyond their standard new-hire onboarding security content, which all three organizations delivered as a one-hour module at employment commencement.

E. Outcome Measures and Statistical Analysis

The primary outcome was the click-through rate (CTR) per condition per wave, defined as the proportion of participants in a given condition who clicked the simulated phishing link

in a given email. CTR was calculated both overall (aggregating all three lure categories) and disaggregated by lure category. Within-condition changes in CTR from Wave 1 (baseline) to Wave 6 (endpoint) were tested using McNemar's test for paired proportions at the individual participant level. Between-condition comparisons of the Wave 1 to Wave 6 change in CTR were tested using chi-square tests of independence. Individual-level moderator analysis examined the association between time-on-module (above vs. below median across all triggered instances) and the probability of clicking on a subsequent campaign wave, using binary logistic regression controlling for lure category and wave number. All analyses were conducted in R version 4.3, with statistical significance set at $\alpha = 0.05$ (two-tailed). Post-hoc power analysis confirmed that the study was powered at 0.81 to detect a difference of 4 percentage points in CTR between conditions at $\alpha = 0.05$.

IV. RESULTS

A. Baseline Equivalence and Overall Click-Rate Trends

At baseline (Wave 1), overall CTRs were comparable across conditions: EMT 18.4%, ILT 19.1%, CTL 18.9%, with no significant between-condition differences ($\chi^2(2)=0.31$, $p = 0.856$), confirming that randomization achieved baseline equivalence. Table I presents Wave 1 and Wave 6 CTRs with absolute and relative changes for each condition. The EMT condition showed a statistically significant reduction from 18.4% to 13.1% (−5.3 pp, a 28.8% relative reduction, $p = 0.002$). The ILT condition showed a non-significant reduction from 19.1% to 17.6% (−1.5 pp, $p = 0.214$). The CTL condition showed a non-significant reduction from 18.9% to 17.2% (−1.7 pp, $p = 0.389$), a decline attributable to general practice effects from repeated exposure to simulation campaigns rather than to any specific training intervention.

TABLE I CLICK-THROUGH RATES BY CONDITION, WAVE 1 VS. WAVE 6 (ALL LURE CATEGORIES COMBINED)

Condition	W1 CTR%	W6 CTR%	Δ (pp)	p
EMT (n = 428)	18.4	13.1	−5.3*	.002
ILT (n = 426)	19.1	17.6	−1.5	.214
CTL (n = 430)	18.9	17.2	−1.7	.389

EMT = Embedded Micro-Training; ILT = Instructor-Led Training; CTL = Control; W1 = Wave 1; W6 = Wave 6; pp = percentage points. *p < .01. Between-condition comparison of Δ: EMT vs. ILT $\chi^2(1)=6.84$, p=.009; EMT vs. CTL $\chi^2(1)=6.21$, p=.013; ILT vs. CTL $\chi^2(1)=0.04$, p=.841.

B. Lure-Category-Specific Effects in the EMT Condition

Table II disaggregates Wave 1 to Wave 6 CTR changes within the EMT condition by lure category, revealing a striking degree of lure-dependence in training effectiveness. The Fake Software Update (FSU) lure category showed the largest reduction: from 21.7% at baseline to 9.3% at Wave 6 (−12.4 pp, $p < 0.001$), a 57.1% relative reduction. This large effect is interpretable within the dual-process and PMT frameworks: the FSU lure contains multiple visually inspectable warning signs (sender domain mismatch, unexpected attachment extension, branding inconsistency) that the 90-second module explicitly highlighted, and that attentive employees could identify through deliberate cue-checking behavior after training.

The Spoofed Invoice (SI) lure category showed a moderate reduction: from 17.2% to 12.8% (−4.4 pp, $p = 0.038$), a 25.6% relative reduction. This intermediate effect is consistent with the SI lure's mixed cognitive demands: it contains some visually inspectable artifacts (unfamiliar vendor name, domain mismatch) but also exploits familiarity with normal invoice-processing workflows, which may partially override trained cue-recognition schemas.

The Urgent Credential Reset (UCR) lure category showed only a minimal non-significant reduction: from 16.3% to 14.5% (−1.8 pp, $p = 0.412$). Critically, the UCR lure had the highest absolute CTR at Wave 6 (14.5%) across all three lure categories in all three conditions, remaining above the EMT group's Wave 6 FSU rate of 9.3% despite six months of repeated exposure and training in the EMT group. This pattern is consistent with the theoretical prediction that urgency-based lures exploit PMT threat appraisal processes in ways that embedded training is not well positioned to counteract: even employees who intellectually understand the warning signs of a UCR lure may respond automatically when confronted with a message warning that their account will be suspended in two hours.

TABLE II EMT CONDITION: CLICK-THROUGH RATES BY LURE CATEGORY, WAVE 1 VS. WAVE 6

Lure category	W1 CTR%	W6 CTR%	Δ (pp)	p
Fake Software Update	21.7	9.3	−12.4***	<.001
Spoofed Invoice	17.2	12.8	−4.4*	.038
Urgent Credential Reset	16.3	14.5	−1.8	.412

***p < .001; *p < .05; ns = not significant. Lure-category analyses restricted to EMT condition participants.

C. Between-Condition Comparisons

Between-condition comparisons confirmed that the EMT condition's overall CTR reduction (−5.3 pp) was significantly larger than both the ILT condition's reduction (−1.5 pp; $\chi^2(1)=6.84$, $p = 0.009$) and the CTL condition's reduction (−1.7 pp; $\chi^2(1)=6.21$, $p = 0.013$). The ILT and CTL conditions did not differ significantly from each other ($\chi^2(1)=0.04$, $p = 0.841$), indicating that a single baseline instructor-led training session produced no detectable improvement in CTR beyond the general practice effect observable in the untrained control group across six months of repeated campaign exposure. This finding replicates the null instructor-led training effect observed in the large healthcare RCT [6], [7] and suggests that the instructor-led-only model, without ongoing feedback loops, is unlikely to produce meaningful behavioral change against real-world phishing campaigns.

Lure-category-specific between-condition analyses confirmed that the between-group difference in overall CTR was driven almost entirely by the FSU lure category (EMT vs. CTL: $\Delta=-10.4$ pp, $\chi^2(1)=18.7$, $p < 0.001$) and to a lesser extent by the SI category (EMT vs. CTL: $\Delta=-4.1$ pp, $\chi^2(1)=4.12$, $p = 0.042$). For the UCR lure category, no significant between-condition difference was observed at Wave 6 (EMT 14.5% vs. ILT 15.8% vs. CTL 15.6%; $\chi^2(2)=0.28$, $p = 0.869$), confirming that neither form of training produced a meaningful protective effect against urgency-framed lures.

D. Engagement Moderator Analysis

Among the 428 participants in the EMT condition, training modules were triggered a total of 1,847 times across all waves (mean triggers per participant=4.31, SD = 1.96). Mean time-on-module across all triggered instances was 47.2 seconds (SD = 22.6 seconds), well below the 90-second module length, indicating that the majority of module instances were not completed in full. The median time-on-module was 52 seconds. A total of 31.4% of triggered module instances showed a time-on-module below 15 seconds, strongly suggesting disengagement or active avoidance of module content.

Individual-level logistic regression analyzing the probability of clicking on a subsequent campaign wave (any wave after the first triggered training instance) found that participants whose mean time-on-module exceeded the 52-second median showed significantly lower subsequent click likelihood (OR = 0.52, 95% CI [0.31, 0.87], $p = 0.011$) than participants below the median (OR = 0.81, 95% CI [0.54, 1.22], $p = 0.204$), after controlling for lure category and wave number. The significant protective effect was confined to above-median engagers, indicating that the aggregate EMT condition effect observed in the primary outcome analysis is driven disproportionately by the approximately 50% of triggered training instances where employees engaged substantively with module content.

V. DISCUSSION

A. Embedded Micro-Training Is Effective but Lure-Dependent

The primary finding—that embedded micro-training produced a statistically significant and practically meaningful 5.3 percentage point reduction in overall CTR that significantly exceeded reductions in both comparison conditions—offers a more optimistic assessment of embedded training effectiveness than the near-null result reported in the landmark healthcare RCT [6], [7]. Several factors may explain this difference. First, the present study used lure categories specifically modeled on documented ransomware delivery techniques and designed training modules that explicitly highlighted the specific warning signs present in those lures, creating a high degree of alignment between training content and test conditions. Second, the 90-second modules in this study included two mandatory comprehension check interactions that required employees to demonstrate recognition of warning signs before proceeding, potentially reducing the opportunity for disengaged click-through that may dilute training effects in commercial platforms without such gates. Third, the present sample of 1,284 employees across three organizations, while well-powered for the observed effect size, is substantially smaller than the 19,500-employee healthcare study, and the effect size observed here (5.3 pp) would have reached conventional statistical significance in the larger study.

The lure-category analysis is the most theoretically and practically significant finding of the study. The dramatic contrast between the FSU effect (−12.4 pp) and the UCR effect (−1.8 pp, non-significant) within the same training condition, across the same participant pool, over the same study period, provides the clearest evidence yet that training effectiveness is not a property of the training program in isolation but of the interaction between training content and lure type. Lures that rely on visually inspectable artifact cues for their effectiveness—sender domain mismatches, unexpected attachment extensions, branding inconsistencies, are trainable, because the relevant warning signs can be brought to conscious attention and become part of the trained employee's deliberate evaluation routine. Lures that rely on urgency-driven emotional override of deliberate evaluation are substantially more resistant to training, because the override mechanism operates prior to and independently of cue-recognition [10], [12].

B. Implications for Ransomware Defense Strategy

The finding that UCR lures, urgency-based credential-reset emails that are among the most common initial access vectors for ransomware affiliate campaigns, are essentially unresponsive to six months of embedded micro-training exposure has direct and significant implications for organizational ransomware defense strategy. It implies that organizations that rely primarily on simulated phishing plus embedded training as their behavioral defense against this

lure category are providing their employees with training that will not materially reduce their susceptibility to the most prevalent attack vector. This represents a significant misalignment between common defensive practice and the actual effectiveness landscape.

The appropriate response is not to abandon training but to complement it with technical controls that do not depend on the individual employee's ability to override an urgency-induced emotional response. Multi-factor authentication (MFA) is the most directly relevant control: by requiring a second authentication factor that the attacker cannot obtain by credential phishing alone, MFA reduces the attacker's ability to exploit successfully harvested credentials regardless of whether the employee was successfully deceived by the UCR lure. DNS and IP filtering of known phishing delivery infrastructure [15] reduces the probability that UCR lures reach the employee's inbox in the first place. Web-based user profiling and behavioral analytics [16] can identify patterns of repeated high-risk clicking behavior and trigger escalated technical controls or targeted intervention for specific individuals. Together, these layered technical controls provide a defense-in-depth posture against UCR lures that training alone cannot deliver.

C. The Engagement Gap and How to Address It

The finding that nearly one-third of triggered training module instances involved less than 15 seconds of engagement, and that the significant protective effect of embedded training was confined to above-median-engagement instances, identifies the engagement gap as a critical implementation failure mode for embedded training programs. This finding suggests that the aggregate effectiveness of the typical commercial embedded training deployment may be substantially lower than the potential effectiveness achievable with the same content if employees were genuinely engaging with it. Organizations investing in embedded training programs should therefore assess module completion and time-on-module data as primary indicators of program health, alongside CTR trend data.

Several mechanisms could address the engagement gap without requiring fundamental redesign of training content. Mandatory comprehension checks that gate module completion on demonstrated identification of the specific warning signs present in the triggering lure, rather than simply requiring a click-through, would directly address the disengagement-through-rapid-clicking pathway. Brief positive reinforcement for correct identification during non-training simulation exposures (employees who do not click receive a notification confirming that they correctly identified a simulated phishing email) has been shown to maintain engagement across repeated campaign waves [9]. The ethical responsibility of security professionals deploying training programs includes an obligation to evaluate engagement metrics rather than treating mere module completion as evidence of effective training [14].

D. Limitations and Future Research

The present study has several limitations that should inform interpretation and future research design. The six-month study duration captures short-to-medium-term training effects; prior longitudinal research suggests that phishing training effects can decay substantially over periods of six to twelve months, and it is unknown whether the FSU category reduction observed here would be sustained at twelve or twenty-four months without continued training exposure [9]. The three participating organizations, while diverse in sector, are not representative of all organizational contexts; in particular, the study underrepresents very large enterprises and organizations with mature pre-existing security cultures, where ceiling and floor effects on CTR may limit the detectable training effect. The simulated phishing lures, while modeled on documented ransomware delivery characteristics, are likely lower in fidelity than real attacks, which may be personalized to individual recipients using publicly available information and may exploit trusted internal communication channels in ways that simulation platforms cannot replicate. The use of click-through as the primary behavioral outcome, while standard in the field, does not capture the full range of downstream behaviors that determine whether a click translates into a security incident—including whether the employee reports the click to IT, enters credentials on a phishing page, or opens a malicious attachment.

VI. CONCLUSION

This paper presented a six-month, three-arm randomized field experiment across three mid-sized organizations, comparing embedded 90-second micro-training, single-session instructor-led training, and no-training control conditions against simulated phishing campaigns modeled on documented ransomware delivery lures. Embedded micro-training produced a statistically significant 5.3 percentage point reduction in overall click-through rates, a result significantly larger than the reductions in both comparison conditions and substantially larger than effects reported in the largest prior randomized evaluation. However, this effect was strongly lure-dependent: it was concentrated in the Fake Software Update category (−12.4 pp), which relies on visually inspectable artifact cues that training can bring to deliberate attention, and was essentially absent for the Urgent Credential Reset category (−1.8 pp, non-significant), which exploits urgency-driven emotional override of deliberate evaluation. Instructor-led annual training alone produced no significant reduction beyond the general practice effect in the control group. Employee engagement with training content, measured by time-on-module, was a significant moderator of individual-level protective effects, with below-median-engagement participants showing no significant training benefit. These findings together support a defense-in-depth model in which embedded micro-training is deployed as an effective complement to—not a replacement for—technical controls including multi-factor authentication and DNS-level phishing filtering, particularly for urgency-based

ransomware delivery lures against which behavioral training alone is demonstrably insufficient.

ACKNOWLEDGEMENT

The authors acknowledge the Faculty of Natural Sciences, University of Jos, Plateau State, Nigeria, for institutional support.

Declaration of Conflicting Interests

The authors declare no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Use of Artificial Intelligence (AI)-Assisted Technology for Manuscript Preparation

The authors confirm that no AI-assisted technologies were used in the preparation or writing of the manuscript, and no images were altered using AI.

REFERENCES

[1] S. Razaulla, C. Fachkha, C. Markarian, A. Gawanmeh, W. Mansoor, B. C. Fung, and C. Assi, "The age of ransomware: A survey on the evolution, taxonomy, and research directions," IEEE Access, vol. 11, pp. 40698–40723, 2023, doi: 10.1109/ACCESS.2023.3268535.

[2] O. Oz, A. Aris, A. Levi, and A. S. Uluagac, "A survey on ransomware: Evolution, taxonomy, and defense solutions," ACM Comput. Surveys, vol. 54, no. 11s, pp. 1–37, Jan. 2022, doi: 10.1145/3514185.

[3] B. Longtchi, R. M. Rodriguez, L. Al-Shawaf, A. Atyabi, and S. Xu, "Internet-based social engineering attacks, defenses and psychology: A survey," arXiv preprint arXiv: 2203.08302, 2022.

[4] R. Mayer, "Applying the science of learning to multimedia instruction," in Psychol. Learn. Motivation, vol. 54, San Diego, CA: Academic Press, 2011, pp. 77–108, doi: 10.1016/B978-0-12-385527-5.00003-4.

[5] M. Han, H. Zhao, X. Ma, and R. Shi, "Influencing factors of information security behavior among college students based on protection motivation theory: Evidence from China," Front. Public Health, vol. 13, p. 1677024, 2025, doi: 10.3389/fpubh.2025.1677024.

[6] G. Ho, A. Mirian, E. Luo, G. Durumeric, and V. Paxson, "Understanding the efficacy of phishing training in practice," in Proc. IEEE Symp. Security Privacy (SP), San Francisco, CA, USA, 2025, doi: 10.1109/SP61157.2025.00076.

[7] G. Ho, A. Mirian, E. Luo, G. Durumeric, and V. Paxson, "Understanding the efficacy of phishing training in practice," Univ. of Chicago / UCSD Technical Report, 2025.

[8] M. De Bona and F. Paci, "A real world study on employees' susceptibility to phishing attacks," in Proc. 15th Int. Conf. Availability, Reliability and Security (ARES), Karlsruhe, Germany, 2020, pp. 4:1–4:10, doi: 10.1145/3407023.3409179.

[9] N. Marshall, D. Sturman, and J. C. Auton, "Exploring the evidence for email phishing training: A scoping review," Computers & Security, vol. 139, art. 103695, 2024, doi: 10.1016/j.cose.2023.103695.

[10] S. S. Alshammari, B. Soh, and A. Li, "Understanding social engineering victimisation on social networking sites: A comprehensive review of factors influencing user susceptibility to cyber-attacks," Information, vol. 16, no. 2, p. 153, 2025, doi: 10.3390/info16020153.

[11] T. Sommestad and H. Karlzen, "The unpredictability of phishing susceptibility: Results from a repeated measures experiment," J. Cybersecurity, vol. 10, no. 1, p. tyae021, 2024, doi: 10.1093/cybsec/tyae021.

[12] J. Li and A. Y. K. Chua, "Think or respond: Understanding the impact of cognitive appraisals on threat detection and phishing susceptibility," Proc. Assoc. Inf. Sci. Technol., vol. 62, pp. 384–395, 2025, doi: 10.1002/pra2.1264.

[13] C. Gerdenitsch, D. Wurhofer, and M. Tscheligi, "Working conditions and cybersecurity: Time pressure, autonomy and threat appraisal shaping employees' security behavior," Cyberpsychol. J. Psychosoc. Res. Cyberspace, vol. 17, no. 4, 2023, doi: 10.5817/CP2023-4-7.

[14] O. O. Blaise, I. Aaron, U. Alfred, and A. Amusa, "Evaluating the ethical frameworks of information security professionals: A comparative analysis," Asian J. Comput. Sci. Technol., vol. 13, no. 2, pp. 61–66, Nov. 2024, doi: 10.70112/ajcst-2024.13.2.4289.

[15] M. S. Islam, M. Sajjad, M. M. Hasan, and M. S. I. Mazumder, "Phishing attack detecting system using DNS and IP filtering," Asian J. Comput. Sci. Technol., vol. 12, no. 1, pp. 16–20, 2023, doi: 10.51983/ajcst-2023.12.1.3552.

[16] S. Ravichandran and K. L. N. Rao, "Design and development of an advancing web information stockpiling for engraved ontology in user contours," Asian J. Comput. Sci. Technol., vol. 11, no. 2, pp. 11–15, 2022, doi: 10.51983/ajcst-2022.11.2.3379.

[17] M. Auwal and S. Lazarus, "Sociological and criminological research of victimization issues: Preliminary stage and new sphere of cybercrime categorization," J. Digit. Technol. Law, vol. 2, no. 4, pp. 915–942, 2024, doi: 10.21202/jdtl.2024.44.