

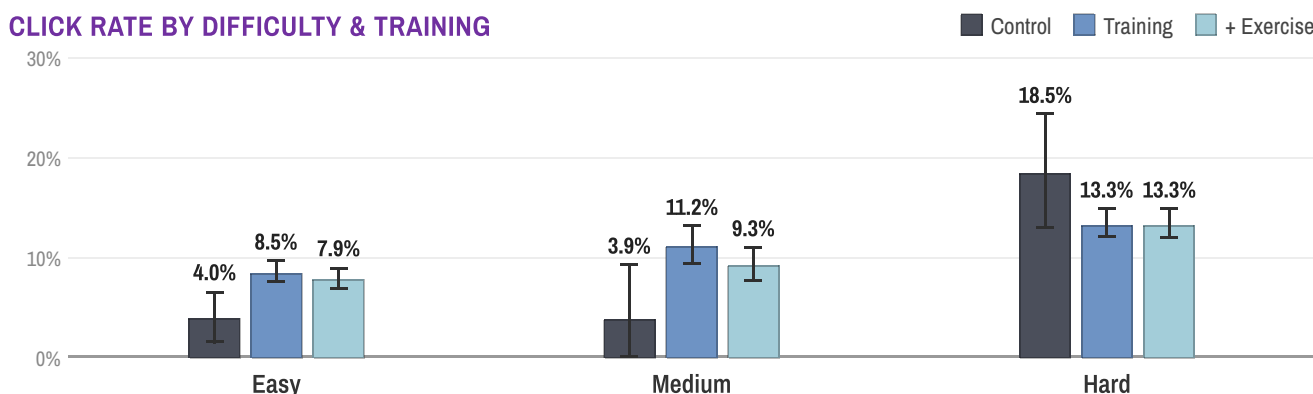


A Reproduction of Phishing Training Inefficacy Grounded in the NIST Phish Scale

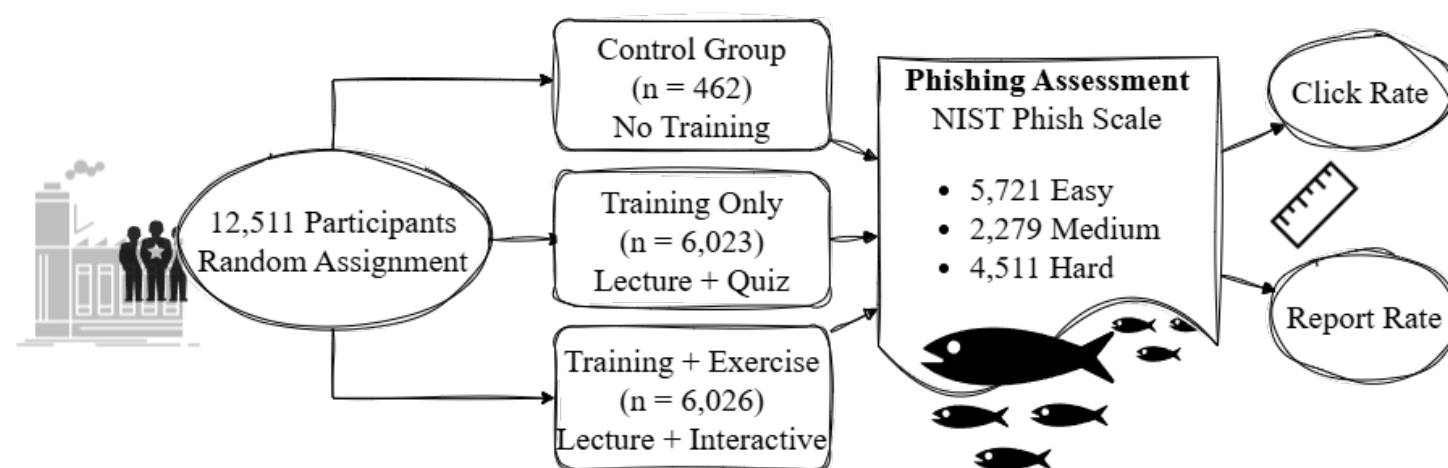
SUMMARY

- ▶ The largest real-world validation of the NIST Phish Scale to date.
- ▶ Difficulty strongly predicts behavior; neither training modality significantly changed clicks ($p=0.450$) or reporting ($p=0.417$) — only **modest protection**, one layer of defense-in-depth.

CLICK RATE BY DIFFICULTY & TRAINING



MATERIALS / METHODS · STUDY DESIGN & KEY FINDINGS



Research Hypotheses:

- ✓ H1: NIST Phish Scale predicts user behavior
- ✗ H2: Training improves detection/reporting
- ✗ H3: Interactive training outperforms lecture-only
- ✗ H4: Training effectiveness varies by difficulty

Key Findings:

- ✓ NIST Scale: 7.0% → 15.0% click rates (easy → hard)
- ✓ Training: No significant main effects ($p > 0.4$)
- ✓ Effect sizes: All below $\eta^2 = 0.01$

Two-factor, between-subjects design (N = 12,511): subjects are randomly assigned a training condition, then receive one phishing lure of NIST-rated difficulty; click and report behavior are measured against five hypotheses.

INTRODUCTION

- ▶ Phishing is a primary attack vector causing breaches, incidents, and financial loss.
- ▶ The Scattered Spider campaigns (2022–23) compromised **9,931 accounts** across 130+ organizations — costing MGM Resorts alone an estimated **\$100M**.
- ▶ Organizations widely mandate anti-phishing training, yet mounting evidence questions its real-world effectiveness.
- ▶ Two gaps remain: **(1)** no standardized measure of lure difficulty; **(2)** few studies compare training modalities at scale.

AIMS / OBJECTIVES

Reproduce phishing-training (in)effectiveness at enterprise scale, in a new context — with standardized difficulty and a modality comparison.

HYPOTHESES TESTED

- H1** — Harder lures → higher click rates.
- H2** — Training improves detection / reporting.
- H3** — Interactive beats lecture for reporting.
- H4** — Training × difficulty interaction.
- H5** — Training improves reporting timeliness.

DISCUSSION

- ▶ After training 12,000+ employees, **no significant main effects** on clicks or reports.
- ▶ Difficulty dominates: orgs can hit low click-rate targets simply by using easy lures — masking true exposure.
- ▶ **52.9%** of campaigns showed “inoculation” (reports preceded clicks) — collective resilience independent of training.
- ▶ Training is just **one layer** of defense-in-depth; AI-generated phishing further erodes the cues the NIST scale relies on.

REFERENCES

1. Steves, Greene & Theofanos. *Categorizing human phishing difficulty: a Phish Scale*. J. Cybersecurity, 2020.
2. Lain, Kostianen & Čapkun. *Phishing in Organizations: A Large-Scale, Long-Term Study (19,000+)*. IEEE S&P '22.
3. Ho et al. *Understanding the Efficacy of Phishing Training in Practice (19,500+)*. IEEE S&P '25.
4. Franz et al. *SoK: Still Plenty of Phish in the Sea — A Taxonomy of Phishing Interventions*. USENIX SOUPS '21.



READ THE FULL PAPER

Scan for the published version & full citations.
doi.org/10.1145/3774904.3792467