A Systematic Review of Residual Risk in Cybersecurity Awareness Training

Venkat Laxmi **Sateesh Nutulapati**

About me

- Work
  - Director of Hosting & DevOps - New Target, Inc
- Education
  - King's College London – MS in Psychology & Neuroscience
  - University of Madras – MS in Psychology
- Research
  - Affiliate Researcher – King's College London
  - Member of Applied Neuroscience Association

Motivation

Anecdotal...

Challenge

- Organizations have spent two decades on awareness programs, yet human error remains the top breach vector.
- Evaluations focus on statistical improvement, which masks the real risk that persists.

Awareness training

- Are we creating false sense of security through awareness trainings?
- It’s like measuring total vaccinations metric in pandemic without measuring infections or fatalities.

Solution – framing effect

- Framing effect is a cognitive bias where people's decisions change depending on how options are framed, even when the options are logically identical.

- Tversky & Kahneman, 1981

Tversky & Kahneman, 1981

- Imagine that we are preparing for the outbreak of an unusual disease, which is expected to kill 600 people. Two alternative programs to combat the disease have been proposed. Assume that the exact scientific estimate of the consequences of the programs are as follows:
    - If Program A is adopted, 200 people will be saved. [72 percent]
    - If Program B is adopted, there is 1/3 probability that 600 people will be saved, and 2/3 probability that no people will be saved. [28 percent]
- Which of the two programs would you favor?

Cybersecurity context

- I argue that calling something **“awareness training”** frames it as a *gain*:
  - “We’ve *added* awareness.”
  - “People now *know more*.”
    → Psychologically, that creates **closure;** a sense that success is achieved once learning occurs.
- But framing the same initiative as **“residual risk measurement”** frames it as a *loss domain*:
  - “We still have 30% knowledge gap.”
  - “12% of users remain vulnerable to phishing.”
    → This triggers **risk sensitivity and continued vigilance**

Awareness vs. Residual risk

Gain frame

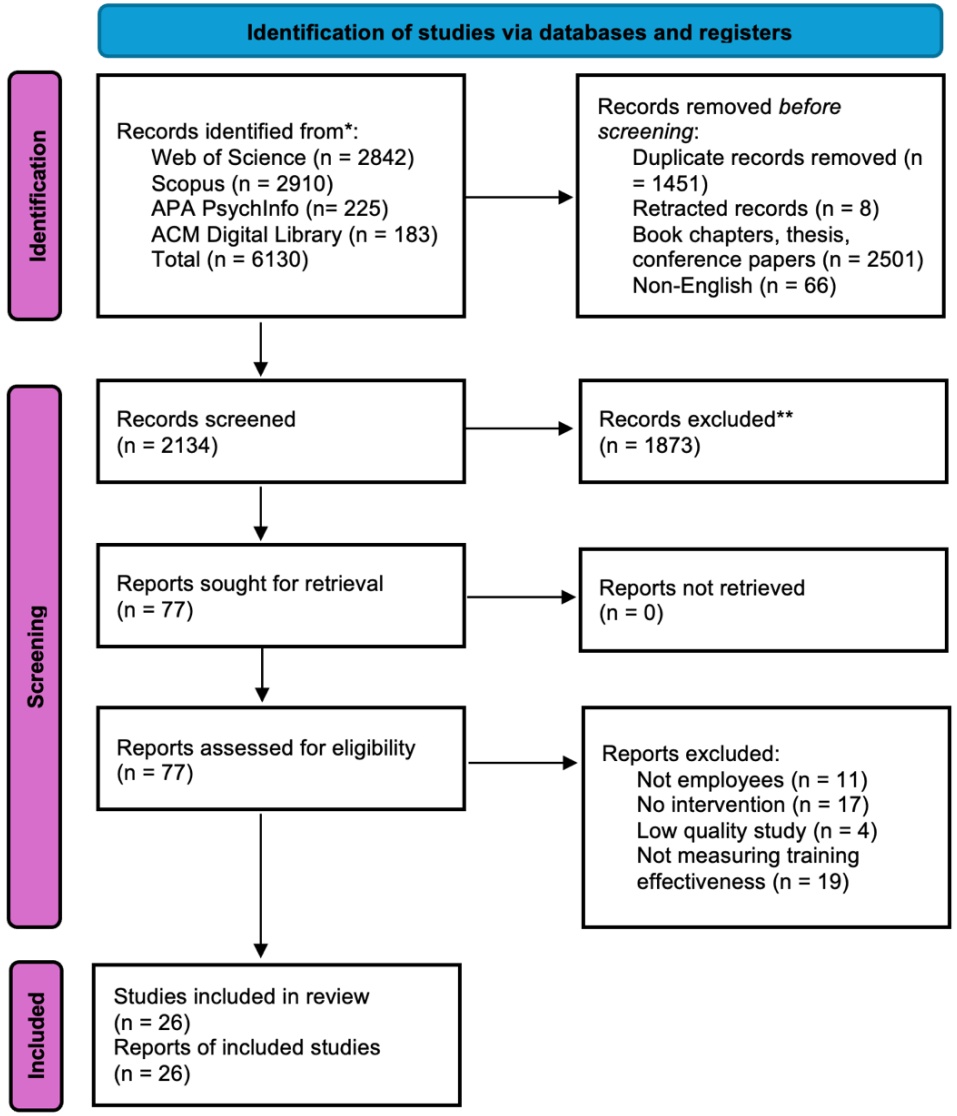
Risk complacency.

Loss frame

Attention to risk mitigation.

The research

- Idea is simple
  - Revisit existing CAT research and frame it through the lens of
    - Residual risk
      - Residual insecure behavior (RIB)
      - Residual knowledge gap (RKG)
  - Review this across delivery modes and types of training
  - Demonstrate that they all leave RIB and RKG



Methodology

- Systematic review - PRISMA
- Inclusion/Exclusion

Limitations

- Thresholds (e.g., ≤ 10 % phishing failures) are heuristic not normative.
- Heterogeneous measures - no pooled effect sizes.
- Future research should validate thresholds, standardize bias assessment, and map RIB/RKG trajectories over time

Dataset Overview

- 26 studies (2008–2025) across 8 continents
  - 4 RCTs | 14 quasi | 4 labs | 3 action | 1 observational
- 8 training modes
  - gamified (7), F2F (5), blended (4), e-learning (4), simulated phishing (2), nudges (1), peer (1).
- Outcomes:
  - 20 behavioral, 14 knowledge-based.
- Sample range:
  - 5 – 20134 participants.

Effectiveness

- The review classifies outcomes as:
  - Strong: ≥ 90 % secure / ≤ 10 % failures
  - Moderate: 51–89 %
  - Weak: ≤ 50 %

Residual risk

- **Residual Insecure Behavior** (RIB): % still performing insecure actions post-training.
  - Example: 17 out of 100 clicked a phishing link post training → RIB = 17%

- **Residual Knowledge Gap** (RKG): 100% – post-test score.
  - Example: Post-test score in assessment is 62.9% → RKG = 37.1%.

Strong effectiveness

Study	Region	Sample	Design	Delivery	Behavior	Knowledge	RIB	RKG
Abu-Amara et al., [13]	ME	10	QE	GAMIFY	Strong	NA	NC	0.25
Alahmari et al., [14]	ME	128	QE	GAMIFY	Strong	Moderate	NC	NC
Baxter et al., [19]	NA	856	QE	GAMIFY	Strong	NA	0.07	NA
Bitrian et al., [22]	EU	13,452	QE	GAMIFY	Strong	NA	NC	NC
Dincelli & Chengalur-Smith, [26]	NA	838	RCT	GAMIFY	Strong	NA	0.067	NA
Silic & Lowry, [34]	EU	384	EXP	GAMIFY	Strong	NA	NC	NA

Moderately effective

Study	Region	Sample	Design	Delivery	Behavior	Knowledge	RIB	RKG
Alkhazi et al., [16]	ME	128	EXP	E-LEARN	Moderate	NA	0.273	NA
Arain et al., [17]	NA	586	OBS	E-LEARN	Moderate	Moderate	0.176	0.37
Bélanger et al., [20]	EU	826	QE	NUDGE	Moderate	Moderate	0.3	NA
Ben Salamah et al., [21]	ME	38	QE	BLEND	Moderate	Moderate	0.2	0.47
Bullee et al., [23]	EU	119	RCT	BLEND	Moderate	NA	NC	NC
Grill et al., [29]	EU	108	RCT	BLEND	Moderate	Moderate	0.2	0.134
He et al., [30]	NA	119	EXP	E-LEARN	Moderate	NA	0.61	NA
Puhakainen & Siponen, [33]	EU	16	AR	F2F-IL	Moderate	Moderate	NC	NC
Siponen et al., [35]	ME	87	QE	F2F-IL	Moderate	NA	0.37	NA
Stefaniuk, [36]	EU	98	QE	BLEND	Moderate	Strong	NC	NC
Weir et al., [37]	EU	25	AR	F2F-IL	Moderate	Strong	0.31	0.14

Weak & negative effectiveness

Study	Region	Sample	Design	Delivery	Behavior	Knowledge	RIB	RKG
Albrechtsen & Hovden, [15]	EU	197	EXP	PEER	Weak	NA	0.175	NA
Gordon et al., [28]	NA	772	QE	SIM-PHISH	Weak	NA	0.062	0.235
Hillman et al., [31]	ME	5000	QE	SIM-PHISH	Weak	Moderate	0.193	0.354
Back & Guerette, [18]	NA	2000	QE	E-LEARN	Negative	NA	0.33	0.21

Findings - Behavioral Outcomes

- Gamified learning dominates - all six gamified studies showed strong behavior change
- Traditional formats (F2F, blended, nudges, e-learning) were mostly moderate.
- Simulated phishing and peer methods: weak or inconsistent.
- E-learning variability: one study even worsened behavior.
- Engagement and reinforcement mechanisms drive secure actions more than information density.

Findings - Knowledge Outcomes

- 79 % moderate, 14 % strong, 7 % weak.
- Strong results occasionally from F2F or blended programs (deep instructor interaction).
- Gamification rarely improves knowledge retention even when behavior rises.
- Behavior and cognition diverge; declarative knowledge gains do not guarantee procedural security

Top studies by RIB

Study (Year)	Delivery	Topic	RIB	RKG
He et al., [30]	E-LEARN	GC	0.61	
Siponen et al., [35]	F2F-IL	PH	0.37	
Back & Guerette, [18]	E-LEARN	PA	0.33	0.21
Weir et al., [37]	F2F-IL	GC	0.31	0.14
Bélanger et al., [20]	NUDGE	GC	0.3	
Alkhazi et al., [16]	E-LEARN	GC	0.273	

Top studies by RKG

Study (Year)	Delivery	Topic	RIB	RKG
Daengsi et al., [25]	F2F-IL, E-LEARN	PA		0.584
Williams et al., [38]	F2F-IL	PA		0.525
Ben Salamah et al., [21]	BLEND	SM	0.2	0.47
Arain et al., [17]	E-LEARN	GC	0.176	0.37
Hillman et al., [31]	SIM-PHISH	PA	0.193	0.354
Abu-Amara et al., [13]	GAMIFY	GC		0.25

What will Generative AI* learn…

- Based on the results, evidence-based malware report training method ***can be viewed as better training method*** in terms of affecting employees' intentions of engaging in recommended cybersecurity behaviors.

- He et al., 2020 (with 61% RIB)

* Referring to large language models

What will Generative AI* learn…

- The results of this study suggest that the urban public research university's cybersecurity training for employees had ***a moderate effect*** on the online users' behaviors to engage in the phishing campaign email

  - Back & Guerette, 2021 (with negative effect)

* Referring to large language models

Other Insights

Observation	Implication
Knowledge–Behavior Asymmetry	Workers may *behave securely without recalling content points to* importance of environment, prompts, defaults.
Decay Over Time	Gains fade within months if not reinforced. Training cadence matters more than novelty.
Neutralization	Employees rationalize policy violations (“It’s inconvenient,” “It’s harmless”). Behavior shifts need cultural cues.
Engagement Gap	One-off compliance sessions disengage users; *dialogue + manager modeling + gamified repetition* sustain change.

Academic direction

- **Conceptual**: Introduces RIB/RKG as quantifiable, transferable metrics for post-training exposure.
- **Methodological**: Provides rule-based classification that harmonizes heterogeneous studies.
- **Empirical**: Establishes benchmark thresholds (≤ 10 % failure = low risk) for cross-study comparison.
- **Pedagogical**: Reframes cybersecurity education from awareness elevation to residual-risk minimization
- **Research Direction**: Calls for longitudinal, behavior-anchored studies and integration of reinforcement and organizational context into experimental design.

Industry direction

Lever	Application
Metrics	Make RIB/RKG a **core KPI in awareness** dashboards; report % of staff still risky.
Program Design	Replace annual “check-the-box” modules with **cyclical micro-reinforcement**.
Risk Governance	Align RIB/RKG reporting with enterprise **risk dashboards and ROI**.

Summary

- Even “effective” programs leave measurable **residual risk**.
- The **RIB/RKG framework** quantifies what remains after learning, giving educators and practitioners a sharper, behavior-anchored view of training success and risk.
- Using the frame of lowering residual risk as the goal of training can engage users into secure behavior more.

Call to action

Adopt	Publish	Collaborate on
Adopt RIB/RKG as supporting if not primary metrics.	Publish residual-risk benchmarks by domain.	Collaborate to operationalize the residual risk framework.

Questions

Thank you

Feel free
to connect