DESIGN DOCUMENT · INSTRUCTIONAL DESIGN PORTFOLIO

Before You Click: An Adaptive Phishing-Judgment Simulation

The design rationale, adaptive model, assessment logic, and measurement plan behind a single-file, evidence-based security-awareness course.

Author: Christopher Kokoski · **Role:** Instructional design, learning engineering, writing, and development

Artifact: Self-contained HTML course (no LMS required) · **Live:** ckokoski.github.io/courses/before-you-click/

Length: 6 to 14 minutes, adaptive · **Status:** Self-directed portfolio build. Meridian Health and all senders are fictional.

1. Executive Summary

Before You Click is an adaptive phishing-judgment simulation. A learner triages a realistic inbox, and the course routes each person down a personal path: a six-email diagnostic maps three competencies, mastery-based routing skips what the learner has proven and remediates what they have not, a boss level moves from detection to response, and a debrief reports not just accuracy but calibration, the alignment between confidence and correctness.

The build demonstrates four capabilities in one artifact: adaptive, mastery-based course design; assessment science (confidence-based marking and calibration reporting); evidence-based multimedia design (animated worked examples in place of instructional text); and measurement transparency (a glass-box panel that shows the learner exactly how the engine routed them, mapped to an xAPI plan).

> **Design stance:** the research is unambiguous that passive video and quiz training does not reduce phishing click rates. This course is built instead on the mechanisms with evidence behind them: active practice with real consequence, immediate contextual feedback, adaptive difficulty, a practiced reporting action, and confidence calibration.

2. Audience, Context, and Constraints

The course targets frontline staff at a fictional regional health system, Meridian Health, consistent with the frontline-healthcare learner persona used across this portfolio: time-poor, interrupted often, on shared or personal devices, and allergic to lecture. Healthcare is the deliberate setting because it carries the highest ransomware stakes, which the cold open dramatizes without resorting to fear alone.

Constraints were chosen to prove craft: one self-contained HTML file with no LMS, no backend, no media assets, and no build step. Sound is synthesized in the browser. The cover illustration is embedded in the file. The whole course is roughly 200 KB and runs identically from a double-clicked file, a static host, or an LMS content window.

3. The Evidence Base

The design followed a research pass across three bodies of work, and the findings drove every major decision.

What does not work

A large 2025 field study (roughly 19,500 employees over eight months, published at IEEE Security & Privacy) found that annual video-and-quiz training produced no significant reduction in phishing clicks,

and that most employees engaged with embedded training for under a minute. Knowledge measures do not predict clicking behavior; production polish does not interrupt forgetting.

What does work

- **Active practice with consequence.** Simulation-based programs with immediate, contextual feedback reduce click rates where passive content does not.
- **A practiced reporting action.** Reporting is muscle memory; programs that drill the physical report action see multiplied report rates, the single strongest program metric.
- **Adaptive difficulty and personalization.** Routing by demonstrated skill outperforms one-size-fits-all delivery.
- **Calibration.** Security research repeatedly finds the riskiest clicks come from people who are confident and wrong, so confidence deserves first-class measurement.
- **Positive framing.** Shame suppresses reporting; celebration and non-punitive feedback sustain it.

Learning science applied

- **Worked examples and cognitive modeling** (Sweller; Bandura): an expert performs the task while thinking aloud, then hands control to the learner.
- **Contrasting cases:** wrong moves shown immediately beside right moves build a contrastive schema that either alone does not.
- **Mayer's multimedia principles:** redundancy (never narrate what the demo shows), signaling (badges and spotlights over prose), segmenting (user-paced, replayable demos), and temporal contiguity (thoughts appear as the cursor acts).
- **Text discipline:** reading sizes at 15px, a 66-character measure, and instructional prose cut by roughly 60 percent in favor of demonstration.

4. Course Architecture

PHASE	WHAT HAPPENS	DESIGN PURPOSE
Cold open	An urgent payroll phish lands at 8:47 AM. The learner clicks or pauses; either way, a breach reconstruction follows, ending at $1.9M and three weeks of downtime.	Consequence-first motivation (Protection Motivation Theory) without shame; the click-or-pause choice is logged, not graded.
Diagnostic	Six emails, two per competency. Verdict (legit or phish) plus a confidence call on every item. No coaching that reveals answers.	Clean baseline that drives routing; confidence data feeds calibration.
Playbook	The cold-open email replayed both ways: three wrong moves with consequences, four calm right moves.	One up-front contrastive model, placed after the diagnostic so it cannot contaminate the assessment.
Training modules (0 to 3)	Each assigned module opens with an animated worked example over the learner's own diagnostic email, then a trimmed rule, then a two-item mastery check with one retry on a miss.	Show, then hand off; remediation only where the diagnostic found a gap.
Boss level	Two response scenarios for everyone: a spear phish built from public information (correct: report) and a plausible internal request for sensitive data (correct: verify out of band).	Moves the skill from detection to response, and teaches calibrated escalation.
Debrief	Path map, calibration report, behavior stats, a personalized retraining plan, badges, and a downloadable result card. A glass-box panel exposes the routing rules and the learner's own data.	Measurement transparency; the Level 2 to Level 3 bridge.

Item-to-competency map

COMPETENCY	DIAGNOSTIC ITEMS	MASTERY CHECKS	CORE DISCRIMINATIONS
Sender forensics	D1 (lookalike IT domain, phish), D2 (real internal newsletter, legit)	M1a, M1b, retry M1r	Display name vs. registered domain; added words, swapped letters, freemail executives, abbreviation domains.
Links and attachments	D3 (subdomain-costume e-sign request, phish), D4 (expected Zoom reminder, legit)	M2a, M2b, retry M2r	Label vs. destination; reading domains right to left; legitimate subdomains vs. attacker decoration; unexpected attachments.
Pressure and pretext	D5 (CEO gift-card freemail, phish), D6 (urgent but ask-free facilities notice, legit)	M3a, M3b, retry M3r	Urgency, authority, and secrecy as levers; urgency alone is not a verdict; what the email asks is the tell.

Each competency pairs a phish with a legitimate email so that "flag everything" fails. False alarms are scored and named as their own outcome, because over-reporting real internal mail has a cost the course teaches honestly.

5. The Adaptive Model

```
mastery(skill) = both diagnostic items correct AND both answered "I'm sure"
if mastered → module skipped  ·  else → module assigned (teach + 2-item mastery check)
if a check is missed → one extra rep on the same skill, spliced into the path
boss level → everyone; response actions instead of verdicts
calibration = confidence × correctness on every verdict  ·  clicks logged as real-world
outcomes
```

The mastery bar is deliberately strict: a correct answer marked "Not sure" assigns the module, because a guess is not a skill. The engine is small enough to be fully inspectable, and the debrief shows it to the learner verbatim. That is a brand decision as much as a design one: adaptive systems earn trust when they can explain themselves.

Path length varies from roughly 15 interactions (clean sweep) to 26 (full remediation with retries). The debrief quantifies what the engine skipped, so strong performers see their time respected.

6. Show, Don't Tell: The Demonstration Engine

Version 3 replaced most instructional prose with a small demonstration engine (internally, the Director) that performs expert behavior on screen: an animated cursor moves over a real rendered email, one-line thought bubbles appear timed to each action, a spotlight isolates the evidence, verdict badges stamp the tells, and a wrong-vs-right montage contrasts the moves that end badly with the calm two-second habits that do not.

Integrity safeguards

- Demos play only over inert decoy emails; they cannot touch the live scoring handlers, and a trusted-event guard means no synthetic click can ever log a consequence.
- Modeling appears only in the playbook and teach modules, never in the diagnostic or boss level, so assessment stays uncontaminated.
- In positive demonstrations the cursor only hovers or press-holds links; click ripples are reserved for the wrong-move montage, so the demonstration never trains the reflex the course exists to untrain.
- Every demo is user-paced with Replay and Skip; Skip renders the full annotated storyboard instantly.
- Under prefers-reduced-motion the demo renders as a complete static storyboard (thoughts as a list, moves as labeled cards), and the narration is mirrored to an aria-live region in both modes.

7. Assessment and Feedback Design

- **Confidence-based marking.** Every verdict carries an "I'm sure / Not sure" call. The debrief reports the 2x2: calibrated, danger zone (confident and wrong), lucky guesses, and honest gaps.
- **Feedback at the moment of error.** A confident-wrong answer is named in the feedback panel ("Sure, and wrong. That is the danger-zone square on your calibration report."), because the confident-wrong moment is the highest-leverage teaching opportunity in the course.
- **Clicking is a consequence, not a question.** Hovering or press-holding a link reveals its destination safely; a quick click is logged on a visible record with an alarm, mirroring the real cost asymmetry.
- **Reporting is a practiced action.** A report control appears on every email. During drills it nudges without bypassing the verdict; at the boss level it is the correct physical response to the spear phish, and time-to-report is measured. The second boss scenario teaches the discrimination: report the masked stranger, verify the plausible colleague, and if verification fails, report without hesitation. A false alarm costs Security seconds; a missed phish costs weeks.
- **Positive reinforcement, never shame.** First peeks earn a brief affirmation through the drills and fade before the boss level; badges celebrate behaviors (First Peek, Zero Clicks, Calibrated, Right Report); wrong answers get evidence, not scolding.

8. The Personalized Retraining Plan

The debrief closes with a ranked plan built from the learner's actual behavior, in priority order: live clicks first, then danger-zone competencies, then misses on emails the learner never inspected ("blind calls"), then unlocked modules, chronic over-flagging, and boss-level response gaps. A clean run earns a spaced-retrieval recommendation instead. Each plan states what a deployed program would do next: targeted assignments and a booster in roughly 30 days. This is the Kirkpatrick Level 2 to Level 3 bridge made visible to the learner.

9. Accessibility (WCAG 2.1 AA)

- Full keyboard operability, including the peek gesture (focus reveals destinations) and all demo controls.
- Focus moves to each new screen; feedback panels receive focus on open.
- Single polite aria-live region carries feedback, demo narration, and badge announcements, sequenced to avoid overwriting.
- prefers-reduced-motion yields a complete static experience, not a broken animated one.
- Contrast verified against the near-black theme (body text about 10.5:1; secondary text above 4.5:1 after two token adjustments found in review).
- Color never carries meaning alone: every state pairs color with a glyph and a text label.
- Sound is optional, synthesized, and mutable; the preference persists.
- Device-adaptive gesture copy: hover on desktop, press-and-hold on touch, taught accordingly.

10. Measurement Plan (Kirkpatrick and xAPI)

LEVEL	IN THIS ARTIFACT	IN A DEPLOYMENT
L1 Reaction	Completion of the optional retake invitation; result-card downloads.	Two-question pulse (confidence to spot; confidence to report) post-course.
L2 Learning	Accuracy by competency, calibration 2x2, mastery-check outcomes, path data.	Same instrumentation via xAPI to an LRS, per cohort.
L3 Behavior	Behavioral proxies inside the simulation: peek rate, live clicks, report actions, and time-to-report.	Simulated-phish click and report rates at 30/60/90 days; time-to-report trend; danger-zone rate on boosters.
L4 Results	Out of scope for a sample.	Real reported-phish volume to the SOC; incident and near-miss trends.

xAPI statement design

EVENT	VERB	OBJECT / RESULT
Verdict on an email	`responded`	Item id, competency, verdict, correctness, confidence, peeked (bool)
Destination peek	`interacted`	Target (sender or link), item id
Live link click	`experienced`	Consequence event, url, item id
Report action	`reported (custom)`	Item id, phase, correct-context (bool), latency seconds
Module routing	`progressed`	Skill, assigned or skipped, diagnostic evidence
Course completion	`completed`	Tier, accuracy, calibration quadrants, clicks, retraining plan items

11. What a Production Deployment Would Add, and Why It Is Not Here

This artifact is deliberately a single self-contained file, so several production capabilities are documented rather than built:

- **Professional narration and captions.** Recorded voiceover cannot live inside a single portable file at acceptable size, and the evidence says narration is a tone lever, not a behavior lever. A deployment would add optional VO with synchronized captions.
- **Spaced repetition.** Behavior change requires cadence: monthly micro-simulations and a 30-day booster targeted by each learner's retraining plan. A single session cannot substitute for a program.
- **LRS integration and dashboards.** The xAPI plan above, wired to an LRS, with cohort dashboards for click rate, report rate, time-to-report, and danger-zone rate.
- **Threat-refreshed content.** Quarterly item refresh from current lures (QR codes, deepfake voice pretexts, MFA fatigue).
- **Localization and org-specific reporting flow.** The report control would bind to the organization's actual button and process.

> **The judgment behind the scope:** a portfolio artifact should prove design and engineering craft, and be honest about what only a deployed program can do. Conflating the two is how training gets bought on polish and fails on behavior.

12. Technical Notes

- One HTML file, roughly 200 KB, vanilla JavaScript and CSS, no libraries, no build step; Google Fonts is the only external request, with system-font fallbacks.
- All sound synthesized with the Web Audio API (no audio files); the cover illustration is embedded as a data URI with an inline SVG fallback.
- The demonstration engine is a small step-driven interpreter (cursor, spotlight, thought, badge, montage primitives) with full teardown between screens; scoring state is untouchable by demos by construction.
- Quality passes: a multi-reviewer adversarial audit (learning design, UX, accessibility, and code) surfaced 30 confirmed findings, all resolved; regression-tested across full novice and clean-sweep paths, reduced motion, mobile widths, and mid-demo interruption.
