



WHITE PAPER

SecNav Assessment & Readiness Methodology

A Technical Overview of Simulation-Based Behavioral Measurement and Continuous Readiness Evaluation

Version 1.0 — August 2026

Security Career Navigator (SecNav) | secnavpro.com

DOCUMENT SCOPE & DISCLAIMER

This document describes the theoretical foundations, technical design, and known limitations of the SecNav simulation-based assessment and readiness measurement system. It is intended for review by workforce development professionals, security standards bodies, hiring institutions, and independent researchers who need to evaluate the validity of SecNav credentials for professional use. This document makes no claims of regulatory approval. It presents our methodology transparently so that institutions may draw their own conclusions.

Document Classification: Public White Paper

Target Audience: Workforce Standards & Security Teams

Contact: research@secnavpro.com

DOI: <https://doi.org/10.5281/zenodo.23056360>

Table of Contents

1. Problem Definition.....	1
1.1 The Measurement Gap in Security Credentialing.....	1
1.2 What the Industry Needs.....	1
2. Assessment Model Overview.....	2
2.1 Design Philosophy.....	2
2.2 Model Components.....	2
3. Simulation Design.....	3
3.1 Scenario Architecture.....	3
3.2 Difficulty Tiers.....	3
3.3 Cognitive Phase Structure.....	4
3.4 Rubric Construction.....	4
4. Behavioral Telemetry.....	5
4.1 What Telemetry We Collect.....	5
4.2 What Telemetry Is Not.....	5
5. Skill Inference.....	6
5.1 From Simulation Performance to Verified Skill Level.....	6
5.2 Skill Coverage and Catalog Scope.....	6
6. DAR Methodology (Decision Action Report).....	7
6.1 Purpose.....	7
6.2 Tactical Score.....	7
6.3 Composure Index.....	7
6.3.1 The Integrity Score.....	8
6.4 Decision Velocity Index.....	9
6.5 Verification Tier and Tactical Designation.....	9
7. RRI Methodology (Role Readiness Index).....	12
7.1 Purpose.....	12
7.2 Formula.....	12
7.3 Critical Skill Gates.....	12
7.4 Role Blueprints.....	12
8. Recency & Cognitive Decay Model.....	13
8.1 Theoretical Foundation.....	13
8.2 Stability Calibration.....	13
8.3 Decayed Level Calculation.....	14
8.4 Decay Status Classification.....	15
9. Framework Mappings.....	15
10. Assessment Integrity.....	17
10.1 The Problem of Online Assessment Integrity.....	17
10.2 The Integrity Score.....	17
10.3 Anti-Automation Detection.....	17
10.4 What the Integrity System Does Not Claim.....	17
11. Credential Verification.....	18
11.1 The KMS Signing Protocol.....	18

11.2 OpenBadges 3.0 and W3C Verifiable Credentials.....	18
11.2.1 Public Key Endpoint.....	19
11.3 What the Credential Claims.....	19
12. Validity and Reliability Considerations.....	20
12.1 Construct Validity.....	20
12.2 Reliability.....	20
13. Known Limitations.....	21
13.1 No Predictive Validity Data.....	21
13.2 Decay Constants Are Theoretically Calibrated, Not Empirically Derived.....	21
13.3 LLM Grader Variance Is Not Formally Quantified.....	21
13.4 Anti-Cheat Detection Is Not Adversarially Complete.....	22
13.5 Simulation Coverage Is Incomplete.....	22
13.6 Role Blueprints Are Not Independently Validated.....	22
13.7 Platform Maturity.....	23
14. Planned Validation Research.....	24
14.1 Grader Consistency Study.....	24
14.2 Expert Practitioner Concurrent Validity Study.....	24
14.3 Predictive Validity Cohort Study.....	24
14.4 Decay Constant Calibration Study.....	24
14.5 Parallel Scenario Development for Test-Retest Studies.....	24
14.6 Independent Expert Blueprint Review.....	25
15. Closing Statement.....	25
16. References.....	26
16.1 Cognitive and Learning Science.....	26
16.2 Assessment and Psychometrics.....	26
16.3 Cybersecurity Frameworks.....	26
16.4 Credential and Technical Standards.....	27
16.5 Regulatory and Industry References.....	27

1. Problem Definition

1.1 The Measurement Gap in Security Credentialing

Traditional security credentials are designed primarily to establish knowledge or competency at a point in time. They do not, by themselves, provide a continuous measure of role-specific operational performance. This is not a criticism of their design — it is a description of their scope. A certification answers the question: *has this individual demonstrated defined knowledge competency at a point in time?* It was not designed to answer the question that is increasingly urgent for hiring organizations: *can this person perform this specific role today?*

This creates three practical gaps that point-in-time credentials were not designed to close:

The Recency Gap. Most professional certifications carry validity windows of two to three years. A credential issued thirty-four months ago is treated identically to one issued last week. No adjustment is made for intervening time, continued practice, or changing threat conditions. A substantial body of cognitive research demonstrates that retention changes over time and is influenced by practice and retrieval [1][3][4]. A credential that does not account for elapsed time cannot, by itself, represent current readiness — and was not designed to.

The Theory-Practice Gap. Standardized knowledge exams reliably assess recall and comprehension — the lower tiers of cognitive function. Real security work — incident containment, forensic investigation, stakeholder communication under pressure — consistently demands higher-order cognition: analysis, evaluation, and synthesis under time and adversarial constraints. Traditional multiple-choice examinations are not designed to directly observe many of these capabilities. The correlation between exam performance and operational performance in security roles has not been demonstrated at the industry level.

Differentiation Collapse. As of 2025, approximately 191,000 professionals hold the CISSP credential globally [16]. Hiring managers cannot meaningfully distinguish between candidates at the same certification level. The credential answers the question *"have you demonstrated minimum knowledge competence at some point in the past?"* It does not answer the question that matters to hiring organizations: *"can this person perform this specific role today?"*

1.2 What the Industry Needs

SecNav is designed to address a measurement need that traditional credentialing was not designed to address: continuous, role-specific evidence of applied operational capability.

SecNav is that instrument. Not a certification. A measurement system. It does not compete with what ASIS, ISC², or NIST have built. It produces a different kind of evidence — behavioral, time-indexed, and role-specific — that existing credentials were not designed to produce, and that hiring organizations increasingly need alongside them.

This document describes how that measurement system works, what we believe it demonstrates, and where our evidence remains incomplete.

2. Assessment Model Overview

2.1 Design Philosophy

The SecNav assessment model rests on three foundational principles derived from established cognitive and psychometric science:

Behavioral primacy. Competence should be inferred from observed behavior under realistic conditions, not from self-reported skill levels or recalled knowledge. SecNav assessments present candidates with simulated operational scenarios and measure the decisions they make, the sequence in which they make them, and the structured operational approach they maintain throughout.

Time-awareness. A readiness signal that does not decay is not a readiness signal — it is a historical record. The SecNav model treats verified competence as a time-decaying quantity, continuously recalculated based on elapsed time since last verification, the difficulty level of that verification, and the performance quality achieved.

Role specificity. Readiness is not generic. A candidate's demonstrated capabilities are evaluated not against an abstract notion of security competence, but against the specific skill requirements of a target role. The same simulation performance produces different readiness signals depending on the role to which it is mapped.

2.2 Model Components

The SecNav assessment model consists of five integrated components:

Component	Function
Simulation Engine	Delivers scenario-based assessments across defined difficulty tiers
Behavioral Telemetry Layer	Collects real-time performance and integrity signals
DAR Engine	Converts raw telemetry into a structured, multi-dimensional performance record
RRI Engine	Aggregates DAR outputs into a time-decaying, role-specific readiness index
Credential Layer	Cryptographically signs and publicly issues verified performance artifacts

3. Simulation Design

3.1 Scenario Architecture

SecNav simulations are structured around realistic security incidents drawn from established threat taxonomies. Each scenario is constructed to require a defined sequence of decisions, each mapped to a specific phase of the incident response lifecycle and a specific tier of cognitive complexity.

Scenarios are not branching trivia trees. They are open-ended dialogue-driven engagements in which the candidate interacts with simulated stakeholders, requests information, issues directives, and makes containment decisions in real time. There is no list of predefined correct answers. The evaluation occurs against a rubric of expected outcomes — specific actions, queries, or decisions that a competent operator in the relevant role should take.

3.2 Difficulty Tiers

Three difficulty tiers are defined, each representing progressively higher operational expectations:

Entry. Scenarios represent the first-response phase of relatively contained incidents. Correct actions are those consistent with established SOPs. The evaluation is lenient with respect to efficiency; it focuses primarily on whether critical containment actions were taken. Intended for professionals early in their career path or in transition.

Professional. Scenarios introduce ambiguity, competing priorities, and GRC trade-offs that mirror real operational environments. Efficient resolution is expected, not merely correct resolution. Designed for mid-career practitioners with established operational experience.

Hardcore. Scenarios are adversarially constructed. Threat actors adapt. Stakeholders may provide misleading information. Policy constraints conflict with optimal technical responses. Near-flawless execution is required for full verification. Intended for senior practitioners whose credentials will be relied upon in high-stakes hiring or audit contexts.

3.3 Cognitive Phase Structure

Each simulation is divided into four operational phases, aligned with Bloom's Taxonomy of cognitive complexity:

Phase	Bloom's Alignment	Focus
Triage & Identification	Remembering / Understanding / Applying	Alert evaluation, IOC identification, SOP activation
Confrontation & Analysis	Analyzing	Entity correlation, log investigation, threat actor attribution
Mitigation & Decisions	Evaluating	Containment vector selection, GRC trade-off resolution
Synthesis & Reporting	Creating / Analyzing	Timeline reconstruction, hardened remediation, CISO briefing

This phased structure serves a dual purpose: it ensures that the simulation exercises the full cognitive range required for the role, and it provides the scoring engine with phase-level performance data rather than a single aggregate score.

3.4 Rubric Construction

Every simulation phase contains a rubric of expected outcomes. Each rubric item is assigned:

- A **weight** reflecting its relative importance to successful incident resolution
- A **category** mapping it to a specific control framework (ISO 27001, OWASP, NIST 800-61, ASIS, or SIRA as appropriate to the scenario domain)
- A **criticality flag** indicating whether failure to satisfy this item constitutes a disqualifying gap

Rubric items are evaluated as **MET**, **PARTIAL**, or **MISSED**. This is not binary grading. Partial credit is meaningful: a candidate who initiates the correct containment action but fails to document it correctly has demonstrated partial competence on that rubric item, which is distinct from failing to initiate containment at all.

4. Behavioral Telemetry

4.1 What Telemetry We Collect

SecNav collects real-time behavioral signals throughout each simulation session. These signals are collected by a **client-side JavaScript instrumentation module** embedded in the simulation interface. Values are transmitted to the backend on each turn submission and stored in the per-turn monitoring record. Server-side timing is not used for individual turn measurement. The signals serve two purposes: performance measurement and integrity verification.

Performance signals collected per turn:

- ❖ **Thinking time (t_think):** Duration between the simulation presenting a state and the candidate beginning to type their response. This measures cognitive processing time before composition begins.
- ❖ **Composition time (t_write):** Duration from first keystroke to submission. This measures the fluency and confidence of the response.
- ❖ **Information queries issued:** The number and type of data requests made to the simulation environment before committing to a decision.

Integrity signals collected per session:

- ❖ **Focus loss events:** Instances in which the browser tab loses focus, suggesting reference to external resources.
- ❖ **Focus loss duration:** Total cumulative time spent outside the simulation window.
- ❖ **Clipboard paste events:** Detected paste actions, which may indicate copying from an AI assistant or external reference material.
- ❖ **Session continuity:** Whether the session was completed without interruption, resumed after disconnect, or abandoned.

4.2 What Telemetry Is Not

Telemetry collection is not surveillance. It is a validity mechanism. Assessment environments that cannot detect whether a candidate is independently producing their responses cannot make meaningful claims about what a score represents. Every instrument that claims to measure individual competence must address the question: "are we measuring this individual, or are we measuring their access to assistance?"

The SecNav telemetry layer is our answer to that question. It does not claim to be perfect — see Section 13 — but it is a direct attempt to address a problem that most online assessment systems do not address.

5. Skill Inference

5.1 From Simulation Performance to Verified Skill Level

A single simulation provides evidence of competence on the skills exercised within that scenario. The SecNav skill inference model translates simulation performance into a structured skill-level claim using the following logic:

Each simulation is tagged with the skills it exercises and the proficiency level at which it exercises them. Difficulty tiers correspond to defined proficiency levels on a continuous 0.0–5.0 scale:

Difficulty	Proficiency Level
Entry	1.5
Professional	3.0
Hardcore	5.0

When a candidate completes a simulation with a **FULLY_VERIFIED** or **PARTIALLY_VERIFIED** outcome, their verified proficiency level for each exercised skill is updated. The verified level is set to the proficiency level of the completed simulation tier, not to the score percentage. A candidate who achieves 85% on a Professional-tier simulation is verified at Level 3.0, not at 2.55. The score affects decay stability (see Section 8), but not the base verified level.

5.2 Skill Coverage and Catalog Scope

At Version 1.0, the SecNav simulation catalog does not provide uniform coverage across all NICE Work Roles defined in NIST SP 800-181 [11]. Simulation coverage across target roles is expanding continuously. Where catalog gaps exist for a given role blueprint, readiness calculations reflect conservative lower bounds, ensuring that missing simulation coverage is never mistaken for verified candidate deficiency.

6. DAR Methodology (Decision Action Report)

6.1 Purpose

The Decision Action Report is the primary output artifact of a SecNav simulation. It is not a certificate of completion. It is a structured performance record across four measurable dimensions. Its function is to give employers and auditors a multi-dimensional view of operational capability that a binary pass/fail result cannot provide.

6.2 Tactical Score

The Tactical Score measures whether the candidate achieved the correct operational outcomes.

Formula:

$$Tactical\ Score = \left(\frac{\sum(W_i \times S_i)}{\sum W_i} \right) \times 100 - P_G$$

Where:

- W_i = normalized weight of rubric item i
- S_i = satisfaction rating: 1.0 (MET), 0.5 (PARTIAL), 0.0 (MISSED)
- P_G = safety guardrail penalty deduced from critical policy or operational violations

The safety guardrail penalty is non-negotiable. A candidate who solves the incident by recommending an action that violates policy, causes collateral damage, or breaches an established control incurs a direct tactical penalty regardless of other rubric achievements. In security work, how you solve a problem is as important as whether you solve it.

6.3 Composure Index

The Composure Index measures the degree to which the candidate maintained structured, efficient operational behavior under pressure.

Formula:

$$Composure\ Index = \max(0, (w_T \times E_T + w_I \times E_I + w_F \times F + w_V \times V) \times 100 - Penalties)$$

Where:

- E_T = turn execution efficiency = $\min(1.0, T_{optimal} / T_{actual})$
- E_I = information query efficiency = $1 - \min(1.0, Q_{excess} / (Q_{total} + 1))$
- F = focus integrity factor derived from window focus retention, blur frequency, and clipboard anomalies
- V = normalized decision velocity

Weights (w_T , w_I , w_F , w_V) are dynamically assigned based on simulation type:

Simulation Type	Turn Eff.	Intel Eff.	Focus	Velocity
Tactical	40%	0%	20%	40%
Forensic / Investigation	30%	20%	50%	0%
Dialogue	40%	30%	30%	0%

Weights are scenario-type-specific because the same behavior carries different significance in different operational contexts. Rapid containment matters more in a live tactical scenario than in a careful forensic analysis.

This component is informed by NIST SP 800-50 Rev. 1's emphasis on measuring the impact of learning programs on workforce capabilities and behavioral change [9].

6.3.1 The Integrity Score

The Integrity Score is a **separate real-time metric** from the Composure Index's focus integrity component F , though both draw on related behavioral signals. Where F is a stateless formula computed at report time from session summary counts, the Integrity Score is maintained continuously by the Trust Engine throughout the session — beginning at 1.0 (full trust) and decaying as violations are detected.

The Trust Engine continuously evaluates multiple behavioral integrity signals throughout the session, including:

- **Focus continuity:** Tab switching frequency and cumulative focus loss duration
- **Input integrity:** Clipboard interaction patterns and anomalous input bursts
- **Environment stability:** Mid-session network/IP continuity and hardware profile consistency
- **Tool tampering:** Direct inspection of simulation runtime infrastructure or client-side instrumentation

Penalty weights are scaled by simulation difficulty tier, reflecting that anomalous behavior carries greater evidential weight in higher-stakes verification contexts. The resulting composite trust score (0.0–1.0) is converted to a percentage and serves as the Integrity Score input for Verification Tier determination (Section 6.5).

6.4 Decision Velocity Index

The Decision Velocity Index measures cognitive speed relative to task complexity.

Formula:

$$V_D = \frac{\sum(L_k \times C_k)}{\sum C_k}, \quad VEI = \int(V_D, SLA_{baseline})$$

Where:

- L_k = thinking time in seconds for turn k (derived from client-side `thinkingTimeMs` telemetry, converted server-side)
- C_k = Bloom's complexity weight for the phase at turn k
- V_D = complexity-weighted decision latency
- $SLA_{baseline}$ = scenario-specific latency benchmark based on allocated duration and operational complexity

SecNav uses Bloom's cognitive levels as a conceptual framework for classifying task complexity across simulation phases [2]. The numerical complexity weights used by the Decision Velocity Index are SecNav-defined parameters — they are not values that Bloom's Taxonomy itself provides or scientifically validates. Their assignment reflects our operational judgment about relative cognitive demand, and they require empirical validation against observed candidate performance data before they can be treated as established constants.

6.5 Verification Tier and Tactical Designation

The DAR does not issue a binary pass or fail. It issues a **Verification Tier** and a **Tactical Designation** as two independent assessments of the same performance.

Verification Tier (confidence in the credential):

Tier	Meaning
FULLY_VERIFIED	All score and rubric completion thresholds met at the required difficulty level
PARTIALLY_VERIFIED	Meaningful engagement demonstrated but full verification standard not met
INSUFFICIENT_TELEMETRY	Session data insufficient to support a reliable claim
FORENSICALLY_COMPROMISED	Behavioral integrity signals exceeded thresholds; session data cannot support a reliable individual capability claim

A note on FORENSICALLY_COMPROMISED: This is a technical status code, not a forensic determination. It indicates that behavioral integrity signals collected during the session crossed thresholds that prevent the system from supporting a reliable individual capability claim. It is a confidence floor, not an accusation or a verdict. SecNav makes no assertion that a candidate designated FORENSICALLY_COMPROMISED cheated or received external assistance — it asserts only that the signal quality is insufficient to issue a meaningful performance credential. This distinction is intentional and is consistent with the integrity system's stated scope in Section 10.4.

Minimum score thresholds by difficulty tier:

Metric	Entry	Professional	Hardcore
Tactical Score	≥ 75%	≥ 85%	≥ 95%
Integrity Score	≥ 50%	≥ 80%	≥ 95%
Composure Index	≥ 40%	≥ 70%	≥ 85%
Decision Velocity	—	≥ 50%	≥ 75%

Rubric Completion Gate. In addition to score thresholds, a rubric weight gate must be satisfied for FULLY_VERIFIED status. The total weight of MISSED rubric items must not exceed:

Difficulty	Max Missed Rubric Weight
Entry	0.15
Professional	0.10
Hardcore	0.0625

A candidate who meets all score thresholds but misses a rubric item above this weight ceiling receives PARTIALLY_VERIFIED rather than FULLY_VERIFIED. This gate reflects our operational judgment that high aggregate scores should not mask complete omission of specific critical outcomes. The threshold values are initial calibrations. We expect to refine them as we study how missed-weight distributions correlate with role-level job performance in the planned validity research (Section 14.3).

Partial Completion Fallthrough. A tactical score between 50% and the difficulty minimum threshold also yields PARTIALLY_VERIFIED — recognizing that the candidate demonstrated

meaningful, substantive engagement without achieving full verification standard. A score below 50% yields **INSUFFICIENT_TELEMETRY** or **FORENSICALLY_COMPROMISED** depending on integrity signals.

Tactical Designation (operational outcome):

Designation	Condition
CONTAINMENT_ACHIEVED	Voluntary conclusion with tactical score $\geq 50\%$
REMEDIATION_REQUIRED	Voluntary conclusion with tactical score $< 50\%$
COMPROMISED	Mission terminated by adversarial breach
RESOURCE_EXHAUSTED	Mission terminated by timeout or turn limit
TACTICAL_ABORT	Session abandoned, unrecoverable protocol deviation, or repeated policy violations

TACTICAL_ABORT via protocol violation cascades into zero turn execution efficiency in the Composure Index — a deliberate design decision that ensures attempts to manipulate the simulation environment produce a measurably degraded performance record, not merely a termination event.

The separation of Verification Tier from Tactical Designation is deliberate. An employer reading a DAR receives two independent signals: *"how much do we trust this score?"* and *"what operationally happened?"* These are different questions and should not be collapsed into a single rating.

7. RRI Methodology (Role Readiness Index)

7.1 Purpose

The Role Readiness Index is a single integer between 0 and 100 that expresses a candidate's current alignment with the skill requirements of a specific target role. It is computed from the candidate's verified skill levels across all skills required by the role, adjusted for time decay (Section 8) and penalized for critical skill gaps.

A score of 100 indicates that the candidate's current verified (and not yet significantly decayed) skill levels meet or exceed all requirements of the target role at the required proficiency levels. A score of 0 indicates either complete absence of verified skills or severe critical skill penalties.

7.2 Formula

$$RRI = \left(\frac{\sum \min(1.0, L_{user} / L_{required}) \times W_i}{\sum W_i} \right) \times 100 - \sum P_{critical}$$

Where:

- L_{user} = current decayed verified level for skill i (see Section 8)
- $L_{required}$ = required proficiency level for skill i in the target role blueprint
- W_i = contribution weight of skill i in the role blueprint
- $P_{critical}$ = 10-point penalty for each unverified critical skill

The $\min(1.0, \cdot)$ cap is important: exceeding a requirement on one skill does not compensate for a gap on another. A candidate who is highly proficient in threat intelligence but has never been verified in incident containment is not a well-rounded candidate for most operational roles.

Excess is not transferable.

7.3 Critical Skill Gates

Every role blueprint designates a subset of skills as critical — competencies whose absence cannot be masked by strength elsewhere. For each such skill that remains unverified, a 10-point deduction is applied directly to the final RRI. The critical gate prevents the composite score from obscuring an unverified critical competency.

7.4 Role Blueprints

Role blueprints define the skills required for a given security role, their required proficiency tiers, their contribution weights, and their criticality status. Blueprints are authored by SecNav and reviewed internally against:

- NIST NICE Work Role definitions (SP 800-181) [11]
- ASIS Certified Protection Professional (CPP) competency domains
- ISC² credential domain structures

- Practitioner review from subject matter experts

At Version 1.0, blueprints have not been independently validated against longitudinal job performance data. This is a stated limitation. See Section 13.

8. Recency & Cognitive Decay Model

8.1 Theoretical Foundation

Hermann Ebbinghaus established through systematic experimental research that memory retention decays exponentially over time in the absence of active recall [3]. His forgetting curve is expressed as:

$$R(t) = e^{-t/S}$$

Where $R(t)$ is the retention fraction at time t (days since acquisition), and S is the stability factor representing the resistance of the memory to decay. Ebbinghaus demonstrated that stability increases with repetition and practice [3]. Subsequent research established that stability is also influenced by depth of encoding and active retrieval [4].

The use of an exponential functional form does not imply that professional skill decay follows the Ebbinghaus forgetting curve exactly. It is an operational modeling choice that requires empirical validation against observed skill decay in security professionals.

SecNav applies a modernized instantiation of this model to verified skill levels. The application is not a literal claim about memory physiology — it is a principled proxy for the well-established empirical fact that professional competence degrades without active practice. The specific functional form is borrowed from Ebbinghaus because it is mathematically well-understood, publicly documented, and independently testable. We use it as an approximation, not as a neurological claim.

8.2 Stability Calibration

The stability factor S is not fixed. It is computed as a function of two variables: the difficulty tier of the simulation and the performance score achieved.

$$S = S_{base} \times \left(1 + 0.5 \times \frac{score}{100} \right)$$

Base stability values by difficulty tier:

Difficulty	Base Stability (S_{base})
Entry	30 days
Professional	90 days
Hardcore	180 days

Rationale for tier-based stability: The depth-of-processing hypothesis in cognitive science holds that material processed at greater depth through active retrieval is retained for longer [4]. Furthermore, cognitive skills decay differently than procedural physical skills [1]. An entry-level simulation exercises surface-level application; a hardcore simulation requires synthesis under adversarial pressure, producing deeper encoding. The 30-, 90-, and 180-day base stability parameters are SecNav operational parameters informed by this principle, not experimentally established universal retention periods. This is a stated limitation.

Rationale for score-based stability adjustment: A higher score on a given simulation suggests more fluent and confident command of the material, which cognitive models associate with stronger encoding. The 50% maximum uplift (at a perfect score) extends stability by up to half again its base value.

8.3 Decayed Level Calculation

The current effective proficiency level of a skill decays from its verified level toward a baseline floor:

$$L(t) = L_{baseline} + (L_{verified} - L_{baseline}) \times e^{-t/S}$$

Where:

- $L_{verified}$ = proficiency level at the time of verification
- $L_{baseline}$ = self-assessed or claimed minimum baseline (the floor below which the model does not decay)
- t = days elapsed since verification
- S = stability factor as computed above

8.4 Decay Status Classification

A skill is assigned a status based on its current retention fraction and its relationship to the required level:

Status	Condition
VERIFIED	$R(t) \geq 0.8$ and decayed level meets required level
VERIFIED_LOW	$R(t) \geq 0.8$ but decayed level below required level
DECAYING	$R(t) < 0.8$ but decayed level still meets required level
DECAYING_LOW	$R(t) < 0.8$ and decayed level below required level

The 80% retention threshold for the **DECAYING** boundary is an initial operational parameter, not a hard scientific constant. It represents our operational judgment that a verified skill at 80% of its peak level remains meaningfully representative of demonstrated competence, while below that threshold the contribution should be flagged as aging. This threshold will be subject to revision as empirical data accumulates.

9. Framework Mappings

SecNav does not replace or compete with established security frameworks and credentialing systems. It operates as a behavioral evidence layer that existing frameworks can independently consume.

The table below documents how SecNav components map to established standards:

Standard / Framework	Organization	SecNav Integration Point
NIST SP 800-181 (NICE) [11]	NIST	Role blueprints are authored against NICE Work Role definitions and task statements
NIST SP 800-61r3 [10]	NIST	Tactical Designations (CONTAINMENT_ACHIEVED , RESOURCE_EXHAUSTED , etc.) are aligned with NIST incident response phases
NIST SP 800-50r1 [9]	NIST	Composure Index is aligned with NIST's outcome-based workforce measurement

Standard / Framework	Organization	SecNav Integration Point
		framework, which shifts from training completion metrics toward observable behavioral performance within role-specific contexts (2024, supersedes withdrawn SP 800-16)
ISO/IEC 27001 [7]	ISO	Tactical rubric items map to ISO 27001 control categories (A.16 for incident management, A.12 for operations security, etc.)
OWASP ASVS [12]	OWASP	Application and infrastructure security rubric items reference OWASP verification requirements
MITRE ATT&CK [8]	MITRE	Threat actor behavior in simulations is modeled against documented ATT&CK techniques and tactics
Bloom's Taxonomy [2]	Bloom	Decision Velocity Index and phase structure are explicitly aligned with the six cognitive levels
Ebbinghaus Forgetting Curve [3]	Ebbinghaus	Decay model formula directly implements the Ebbinghaus exponential retention model
ASIS (Industry Certification Domain)	ASIS International	Physical and operational security role blueprints reference ASIS CPP and PSP competency domains. Note: ASIS CPP/PSP are professional certifications, not engineering standards.
W3C Verifiable Credentials [14]	W3C	All issued DAR credentials use the OpenBadges 3.0 / W3C VC standard

Important clarification: These mappings represent our authorial intent and architectural design decisions. They have not been independently validated or endorsed by any of the listed organizations. We make no claim that use of SecNav credentials satisfies any regulatory

requirement. That determination rests with the relevant regulatory authority and the employing organization.

10. Assessment Integrity

10.1 The Problem of Online Assessment Integrity

Any unproctored online assessment faces the fundamental challenge that the assessment environment cannot guarantee the candidate is independently producing their responses. This challenge is particularly acute in the current period, given the wide availability of large language model assistants capable of producing high-quality security responses in seconds.

We treat this as an engineering problem, not a policy problem. Policies that prohibit AI assistance are unenforceable. Technical systems that make AI assistance detectable are not.

10.2 The Integrity Score

The Integrity Score is a continuous 0–100 metric computed from behavioral telemetry signals. It is not a binary cheat-detection flag. It is a confidence measure: *"to what degree do we believe the performance data we collected represents this individual's independent capability?"*

A low Integrity Score does not constitute proof of cheating. It constitutes insufficient evidence to make a confident individual capability claim. The system responds accordingly — downgrading the Verification Tier rather than issuing an accusation.

10.3 Anti-Automation Detection

SecNav uses multiple behavioral signals to identify potential automation, script pasting, and external AI assistance:

- **Response Latency Analysis:** Thinking time (t_{think}) is evaluated relative to the cognitive complexity of the scenario phase. Anomalously fast response times on complex analytical and synthesis tasks trigger integrity flags, indicating potential external retrieval or automated injection.
- **Composition Dynamics:** Typing velocity and composition patterns are continuously monitored during turn entry. Composition rates that significantly exceed normal human typing thresholds indicate paste events or automated input generation, distinguishing natural composition from machine-assisted insertion.

These operational parameters are structured by difficulty tier and cognitive phase. They represent initial operational baselines that are subject to ongoing empirical refinement as population-level latency and composition distributions accumulate.

10.4 What the Integrity System Does Not Claim

The integrity system makes no claim to be adversarially complete. A sufficiently motivated individual who understands the telemetry collection model could potentially engineer telemetry

patterns that pass integrity checks while still relying on external assistance. We do not claim otherwise. What we do claim is that the integrity system raises the cost of gaming the assessment above the equivalent cost of simply studying for and passing a traditional exam. It is not insurmountable. It is a higher bar.

11. Credential Verification

11.1 The KMS Signing Protocol

Every DAR issued upon successful simulation completion is cryptographically signed using a dedicated **Hardware Security Module (HSM) / Secure Cloud Key Management Service (KMS)**. The signing process uses the **RDFC-1.0 canonicalization algorithm** (formerly known as URDNA2015), as specified in the W3C RDF Dataset Canonicalization Recommendation [17]:

1. The proof options document and the credential document are each independently normalized using the RDFC-1.0 algorithm.
2. Each normalized document is hashed with SHA-256.
3. The two SHA-256 digests are concatenated to form the signing payload — this two-document concatenation is the RDFC-1.0 standard's method for binding both the credential content and its proof metadata into a single verifiable unit.
4. The concatenated payload is signed using an asymmetric **Ed25519 key** held in an isolated cryptographic enclave. Private key material is never exposed outside the secure KMS boundary or to application runtime memory.
5. The resulting signature conforms to the W3C Verifiable Credentials Data Integrity specification for EdDSA signatures [15].
6. The encoded signature and proof metadata are embedded in the issued credential's proof section.

When an employer or auditor verifies a DAR credential via the SecNav verification endpoint or by scanning the embedded QR code, the system reconstructs the same RDFC-1.0-canonicalized payload and validates the signature against the public key. Any modification to the credential — including altering a score, a timestamp, or a verification tier — changes the canonicalized document hash and produces an immediate verification failure.

11.2 OpenBadges 3.0 and W3C Verifiable Credentials

Credentials are issued in compliance with the **OpenBadges 3.0** specification [13], which is built on the **W3C Verifiable Credentials Data Model** [14]. This means:

- Credentials are machine-readable and compatible with any W3C VC-aware verification system
- The credential includes a proof section containing the cryptographic signature
- Verification does not require the verifier to rely on the integrity of the displayed credential document itself; the signature and issuer key establish provenance and detect modification

- Credentials can be shared as portable, self-verifying documents to any platform that supports the VC standard

11.2.1 Public Key Endpoint

SecNav publishes the current verification public key at a well-known URI following the RFC 5785 convention:

```
GET https://api.secnnavpro.com/.well-known/kms-public-key
```

Any party can fetch this key and use it to independently verify the cryptographic signature on any SecNav-issued DAR credential, without contacting SecNav. This makes the authenticity and integrity of issued credentials independently auditable by employers, institutions, or standards bodies using standard cryptographic tooling.

This endpoint reflects the currently active signing key. SecNav will update the endpoint on key rotation. Credentials signed under a prior key remain verifiable against the key embedded in their proof section at issuance.

11.3 What the Credential Claims

A SecNav DAR credential claims, specifically:

On [date], candidate [identifier] completed a [difficulty]-tier simulation in the domain of [scenario domain]. The following performance metrics were recorded under telemetry integrity conditions assessed at [Integrity Score]. The performance produced a [Verification Tier] outcome and a [Tactical Designation] tactical designation. This record has not been modified since issuance.

The credential does not claim that the candidate will perform equivalently in a live environment. It does not claim that the simulation perfectly replicates real operational conditions. It claims only that this performance occurred, was measured in this way, and has not been altered.

12. Validity and Reliability Considerations

12.1 Construct Validity

Construct validity addresses whether an assessment measures what it claims to measure [5]. SecNav claims to measure operational readiness through behavioral performance under simulated conditions.

Evidence supporting construct validity:

- The assessment design is grounded in established cognitive frameworks (Bloom's Taxonomy, Ebbinghaus) rather than ad hoc scoring conventions
- Rubric items are mapped to established control frameworks (NIST, ISO, OWASP), providing criterion alignment with recognized professional standards
- The use of behavioral telemetry rather than recalled knowledge directly targets the operational capability construct

Evidence not yet available:

- No predictive validity studies have been conducted relating SecNav scores to real-world job performance outcomes
- No concurrent validity studies have been conducted comparing SecNav RRI scores against expert practitioner assessments of the same individuals
- No discriminant validity studies have been conducted demonstrating that SecNav scores differ meaningfully between novice and experienced practitioners

These are not rhetorical caveats. They represent the genuine current state of the evidence base. We are at Version 1.0. We know what we have not yet proven.

12.2 Reliability

Reliability addresses whether the assessment produces consistent results across administrations.

The LLM grading variance issue: SecNav uses a large language model to evaluate candidate responses against defined rubrics. LLM graders are not perfectly deterministic. The same response submitted twice may receive marginally different rubric evaluations. We mitigate this through rubric anchoring — providing the grader with explicit scoring criteria and reference examples for each rubric item — but inter-rater variance exists. At Version 1.0, the magnitude of this variance has not been formally quantified. This is a significant known limitation.

Test-retest reliability: No formal test-retest reliability studies have been conducted. A candidate completing the same scenario twice would be expected to perform better on the second attempt due to scenario familiarity, introducing confounding. Meaningful test-retest studies would require equivalent-difficulty parallel scenarios. Developing this parallel scenario infrastructure is in the research roadmap.

13. Known Limitations

This section is the most important section of this document for institutional review purposes. It documents the difference between what we have demonstrated and what we are still attempting to prove. We publish it not because we are required to, but because we believe that honest enumeration of limitations is the first requirement of scientific credibility.

13.1 No Predictive Validity Data

The limitation: We have not demonstrated that SecNav scores predict real-world job performance. This is the foundational validity question for any assessment instrument used in hiring [6]. Without this evidence, any claim that a high RRI predicts on-the-job success is unsubstantiated.

Current status: We are tracking early user outcomes with the intention of conducting predictive validity studies once a sufficient cohort of users has been hired into roles using SecNav credentials. We expect meaningful data to become available in 12–24 months at current growth projections.

Implication for institutions: SecNav credentials should currently be treated as one additional data signal, not as a replacement for interviews, technical assessments, or reference checks. We make no claim otherwise.

13.2 Decay Constants Are Theoretically Calibrated, Not Empirically Derived

The limitation: The base stability values (30, 90, and 180 days for Entry, Professional, and Hardcore respectively) are initial operational parameters informed by cognitive science literature, specifically the depth-of-processing framework associated with Ebbinghaus and subsequent researchers. They are not derived from empirical measurement of actual skill decay in security professionals.

What this means in practice: The decay model is theoretically sound, but the specific threshold values are our best current calibration, not validated constants. A professional's actual retention of a hardcore-level skill may decay faster or slower than 180 days depending on their subsequent work experience, domain adjacency, and individual cognitive factors not captured by the model.

Planned resolution: As the platform accumulates longitudinal data on user re-verification patterns and score trajectories, we intend to refine the decay constants against actual observed decay rates. This is explicitly in our research roadmap.

13.3 LLM Grader Variance Is Not Formally Quantified

The limitation: The rubric evaluation engine uses a large language model grader. We have implemented prompt engineering controls and rubric anchoring to reduce variance, but we have

not formally measured inter-evaluation consistency — the equivalent of inter-rater reliability in human-scored assessments.

What this means in practice: A candidate who submits a response that sits near a rubric item's MET/PARTIALLY_MET boundary may receive different evaluations across grading runs. The magnitude of this uncertainty is unknown at Version 1.0.

Planned resolution: We intend to conduct a formal grader consistency study using a held-out set of responses with known ground-truth expert ratings. This will allow us to quantify grader variance and express it as a confidence interval on DAR scores.

13.4 Anti-Cheat Detection Is Not Adversarially Complete

The limitation: The behavioral integrity system is designed to detect casual use of AI assistants and scripted responses. It is not designed to defeat a sophisticated adversary who has studied the telemetry collection mechanism. Someone who deliberately types responses at human speeds, introduces artificial pauses, and avoids focus loss events while still sourcing answers from an AI could potentially produce an inflated Integrity Score.

What this means in practice: The Integrity Score represents our current best estimate of assessment validity, not a mathematical guarantee. High Integrity Scores on assessments with low inherent difficulty may be more susceptible to gaming than high Integrity Scores on Hardcore-tier simulations, where the quality of reasoning required is harder to fake convincingly even with AI assistance.

13.5 Simulation Coverage Is Incomplete

The limitation: At Version 1.0, the SecNav simulation catalog does not cover all skills required by all security roles in the platform's role blueprints. Where coverage gaps exist, RRI scores for roles dependent on those skills represent conservative lower bounds rather than definitive ceiling estimates. Coverage gaps are a function of catalog growth and are addressed through continuous scenario authoring.

13.6 Role Blueprints Are Not Independently Validated

The limitation: The skill requirements, weights, and criticality designations in SecNav role blueprints have been authored against recognized engineering frameworks (NIST NICE) and established industry certification domains (ASIS, ISC²) and reviewed internally. They have not been validated through independent expert panel review or empirical job task analysis. A formally validated job task analysis is the gold standard for this work; SecNav has not yet conducted one.

Implication: The role blueprints represent our best current model of what each role requires. They should be treated as a working model subject to revision, not as a finalized authoritative job specification.

13.7 Platform Maturity

The limitation: SecNav launched in August 2026. At the time of this document's publication, the platform has a small user base and limited longitudinal data. Many of the validity claims made in this document rest on the strength of the theoretical framework rather than on platform-derived empirical evidence.

We are a new measurement instrument. We have been designed carefully and grounded in established science. We have not yet been proven at scale. We intend to be.

14. Planned Validation Research

The following research activities are planned to address the limitations documented in Section 13. We publish this roadmap to allow institutions to assess what evidence will be available at what time frame.

14.1 Grader Consistency Study

Objective: Quantify inter-evaluation variance of the LLM rubric grader

Method: Collect a held-out set of candidate responses. Have the grader evaluate each response 20 times. Compute coefficient of variation per rubric item. Compare against human expert ratings for a subset.

Timeline: Target Q1 2027

14.2 Expert Practitioner Concurrent Validity Study

Objective: Determine whether SecNav scores correlate with expert assessments of the same individuals

Method: Recruit practicing security professionals with verifiable experience. Have them complete SecNav simulations. Collect expert peer ratings of their operational capability. Compute correlation between RRI and peer ratings.

Timeline: Target Q2 2027 pending user base growth

14.3 Predictive Validity Cohort Study

Objective: Determine whether SecNav RRI scores predict job performance in security roles

Method: Track users who are hired into security roles following SecNav-assisted recruiting. Collect manager performance ratings at 6 and 12 months. Correlate against RRI at time of hire.

Timeline: Target Q4 2027; requires sufficient hiring cohort

14.4 Decay Constant Calibration Study

Objective: Replace theoretically-derived decay constants with empirically-validated values

Method: Analyze longitudinal re-verification data. For each difficulty tier, fit observed score trajectories to the Ebbinghaus exponential model to derive empirical stability parameters.

Timeline: Ongoing; first meaningful analysis targeted Q2 2027

14.5 Parallel Scenario Development for Test-Retest Studies

Objective: Enable formal test-retest reliability measurement

Method: Develop parallel scenario sets covering the same skills and cognitive demands at equivalent difficulty, with different surface content.

Timeline: Target Q3 2027

14.6 Independent Expert Blueprint Review

Objective: Subject role blueprints to independent peer review

Method: Convene a panel of senior practitioners in each covered role domain. Provide them with current blueprints. Collect ratings of skill relevance, weight appropriateness, and criticality designations. Revise accordingly.

Timeline: Target Q2 2027

15. Closing Statement

SecNav is a measurement instrument. Like any measurement instrument, its value is determined not by the ambitions of its designers but by the quality of the evidence it accumulates over time.

We have built a system grounded in established cognitive science, mapped to recognized professional frameworks, designed with integrity safeguards, and issued via cryptographically verifiable credentials. We believe these design choices are correct. We have not yet proven all of them at the scale and rigor that would satisfy the highest standards of psychometric validation.

We know the difference between what we have demonstrated and what we are still trying to prove. This document is our commitment to proving it.

16. References

16.1 Cognitive and Learning Science

1. **Arthur, W. Jr., Bennett, W. Jr., Stanush, P. L., & McNelly, T. L. (1998).** Factors that influence skill decay and retention: A quantitative review and analysis. *Human Performance*, 11(1), 57–101. https://doi.org/10.1207/s15327043hup1101_3
2. **Bloom, B. S. (Ed.). (1956).** *Taxonomy of Educational Objectives: The Classification of Educational Goals. Handbook I: Cognitive Domain*. New York: David McKay Company.
3. **Ebbinghaus, H. (1885).** *Memory: A Contribution to Experimental Psychology*. (H. A. Ruger & C. E. Bussenius, Trans.). Teachers College, Columbia University.
4. **Roediger, H. L., & Karpicke, J. D. (2006).** Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, 17(3), 249–255.

16.2 Assessment and Psychometrics

5. **American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014).** *Standards for Educational and Psychological Testing*. Washington, DC: AERA.
6. **Kane, M. T. (2006).** Validation. In R. L. Brennan (Ed.), *Educational Measurement* (4th ed., pp. 17-64). Westport, CT: Praeger.

16.3 Cybersecurity Frameworks

7. **International Organization for Standardization. (2022).** *Information security, cybersecurity and privacy protection — Information security controls (ISO/IEC 27001:2022)*. Annex A.
8. **MITRE Corporation. (2024).** *MITRE ATT&CK® Framework*. Enterprise Matrix. <https://attack.mitre.org/>
9. **National Institute of Standards and Technology. (2024).** *Building a Cybersecurity and Privacy Learning Program (NIST SP 800-50 Rev. 1)*. <https://doi.org/10.6028/NIST.SP.800-50r1>
10. **National Institute of Standards and Technology. (2025).** *Incident Response Recommendations and Considerations for Cybersecurity Risk Management: A CSF 2.0 Community Profile (NIST SP 800-61r3)*. <https://doi.org/10.6028/NIST.SP.800-61r3>
11. **National Institute of Standards and Technology. (2020).** *Workforce Framework for Cybersecurity (NICE Framework) (NIST SP 800-181 Rev. 1)*. <https://doi.org/10.6028/NIST.SP.800-181r1>
12. **OWASP Foundation. (2021).** *OWASP Application Security Verification Standard (Version 4.0.3)*. <https://owasp.org/www-project-application-security-verification-standard/>

16.4 Credential and Technical Standards

13. **1EdTech Consortium. (2023).** *Open Badges Specification v3.0*.
<https://www.imsglobal.org/spec/ob/v3p0>
14. **World Wide Web Consortium. (2023).** *Verifiable Credentials Data Model v2.0*. W3C Recommendation. <https://www.w3.org/TR/vc-data-model-2.0/>
15. **World Wide Web Consortium. (2024).** *Data Integrity EdDSA Cryptosuites v1.0*. W3C Recommendation. <https://www.w3.org/TR/vc-di-eddsa/>

16.5 Regulatory and Industry References

16. **ISC2. (2025).** *ISC2 Cybersecurity Workforce Study 2025*. <https://www.isc2.org/research>
17. **W3C RDF Dataset Canonicalization and Hash Working Group. (2024).** *RDF Dataset Canonicalization*. W3C Recommendation. <https://www.w3.org/TR/rdf-canon/>

SecNav — Security Career Navigator

Assessment & Readiness Methodology | Technical White Paper

Canonical URI: secnavpro.com

Permanent DOI: <https://doi.org/10.5281/zenodo.23056360>

Version 1.0 • August 2026 • Distribution: Public (Unrestricted)