

RESEARCH PAPER

How Verification Should Scale

A proposal for varying CMMC Level 2 review depth while preserving safeguards, human authority, escalation, and Government oversight.

STATUS	LAST REVIEWED	RESEARCH LANES	CLAIM REGISTER
Current	August 13, 2026	Mechanism · Evaluation · Principles · Analysis	C-01 · C-03 · C-04 · C-07 · C-08

Current CMMC 20X verification design as of August 13, 2026. The routes and tests are proposals for Government evaluation; they are not current CMMC assessment types, eligibility rules, legal advice, or measured outcomes.

The policy question is not whether verification matters. It is: **how should review depth increase as the consequence of a wrong conclusion increases?**

One uniform route is easy to describe. It can also spend scarce independent capacity on cases with very different missions, data, exposure, and change. Pure self-review has the opposite weakness: it asks the same organization to implement, interpret, and judge its own safeguards even where failure could have severe consequences.

CMMC 20X proposes graduated Level 2 verification. Every organization remains responsible for the complete near-term Level 2 baseline. What changes is the depth, independence, and frequency of review. This is a design proposal. The current CMMC program defines its own assessment types and legal effects under 32 CFR part 170.

Scale scrutiny, keep the baseline

Graduation must never become an implicit reduction in safeguards. Each route addresses the same requirements and applicable assessment objectives. Each produces a traceable record. Each preserves contractor responsibility and authorized human decisions.

PROPOSED ROUTE	INTENDED POLICY USE	ESSENTIAL CONTROLS
Evidence-backed self-review	Lower-consequence Level 2 work with support that Government can sample	Organizational affirmation, structured evidence, change triggers, random and risk-based sampling, mandatory escalation
Assisted human review	Cases needing qualified review short of a complete independent assessment	Qualified reviewer, explicit methods, software/human separation, conflict disposition, appeal and oversight
Independent C3PAO assessment	Highest-consequence missions, sensitive data, significant unresolved risk, or conditions requiring full independent examination	Authorized assessor, required independence, complete assessment procedures, quality oversight, formal findings

These labels are placeholders for testing. Government would need to define who qualifies, which contracts are eligible, which procedures apply, how results are recorded, and what legal or acquisition effect follows.

Route selection starts with consequence

The primary inputs should describe the consequence and exposure of the work:

- mission effect if the supplier or information is compromised;
- sensitivity, aggregation, and operational value of the CUI;
- threat exposure, including access paths and known targeting;
- supplier criticality, substitutability, and dependencies;
- architecture and concentration risk, including shared providers; and
- material changes that alter the prior risk picture.

These factors require Government and acquisition context. Contractors know their systems. Primes may know program dependencies and flowdown context. Government owns mission consequence and the authority to assign a verification route.

Evidence quality is a different signal. Missing, stale, conflicting, or untraceable support can make a case unready or trigger escalation. A complete package cannot make a high-consequence mission low risk. A disorganized package cannot by itself prove that the mission is critical. Keeping these inputs separate prevents presentation quality from becoming a proxy for national security consequence.

Review depth has several dimensions

Scaling verification is more precise than choosing among three labels. Policy can vary several dimensions:

Independence. Who employs, selects, and oversees the reviewer? Which conflicts are prohibited?

Coverage. Are all objectives examined? Which artifacts, populations, locations, and periods are included?

Method. What combination of examine, interview, and test procedures is required? Which sampling rules apply?

Frequency. When does the review recur, and which events reopen the result?

Corroboration. Which claims require direct source access, observation, independent testing, or external data?

Oversight. How often does Government sample results, review reviewer quality, or repeat work independently?

Consequence. What status, eligibility, correction period, appeal path, or enforcement action can follow?

A mature policy should specify these dimensions for each route. Otherwise “assisted review” can mean anything from file validation to a de facto full assessment.

Software supports scale through repeatable work

Software can collect source records, normalize formats, check required fields, compare populations, detect stale evidence, follow references, identify known conflicts, and suggest objective mappings. AI may help summarize complex material or propose questions.

The authority boundary must remain visible. A deterministic check records the rule it ran and the result. An AI output records its model or system version, inputs, operation, and limitations. A qualified person reviews the evidence, resolves exceptions, and authors the finding. An authorized organization makes the certification, affirmation, acquisition, acceptance, or enforcement decision.

This separation is especially important in lower-intensity routes. Automation can expand review reach only if users can inspect when it failed, what it missed, and whether a person relied on it.

Escalation is part of every route

A route cannot be fixed for the life of a contract. Escalation conditions should be public and testable. Examples include:

- evidence conflicts that affect a requirement or claimed population;
- suspicious provenance, tampering, or misleading embedded instructions;
- material boundary, provider, architecture, or CUI-flow change;
- a significant incident, failed security test, or repeated corrective-action failure;
- inability to reconcile expected and observed assets or accounts;
- conditions outside the authorized method or reviewer competence;
- Government sampling that disagrees materially with the original result; and
- new mission or threat information that raises the consequence of failure.

Escalation can take several forms: targeted collection, additional human review, expanded procedures, Government examination, or full independent assessment. The record should state the trigger, decision maker, new scope, and resolution.

De-escalation needs equal discipline. It should follow defined criteria and an authorized decision. A vendor score, a clean dashboard, or a structurally valid package should never reduce review depth on its own.

Sampling makes distributed verification credible

Evidence-backed self-review depends on meaningful sampling. Random selection estimates general quality and discourages strategic selection of easy cases. Risk-based selection focuses attention on critical missions, shared providers, reported incidents, unusual evidence patterns, and prior problems. Complaint and referral channels catch cases that quantitative criteria miss.

Sampling should test both contractor implementation and route performance. Did the organization support its claims? Did the assigned method detect material conflicts? Did reviewers apply procedures consistently? Did the escalation rule work?

Results should feed policy adjustment. If one route repeatedly misses a class of identity, provider, or scoping failure, Government can change procedures, raise sampling, narrow eligibility, or move those cases to deeper review.

Failure modes a scalable system must expose

Risk-score theater. A composite number hides subjective inputs and sharp threshold effects. Publish the factors, sources, decision rights, and reasons for route assignment. Permit review and correction of factual inputs.

Supplier gaming. Organizations restructure systems or descriptions to seek a lighter route. Use mission and data context, provider concentration, and Government-held information. Audit changes near thresholds.

Reviewer drift. Assisted reviewers gradually apply different standards. Use versioned procedures, calibration cases, disagreement review, quality sampling, and retraining.

Automation bias. Reviewers accept a machine suggestion without examining contrary evidence. Preserve source links, require dispositions for material conflicts, and measure unsupported agreement with software.

Capacity displacement. New review work consumes the same qualified people needed for full assessments. Measure labor by role and case type. Do not assume that a new route creates capacity.

Unequal access. Small suppliers face tools or expertise requirements they cannot meet. Keep the evidence profile open, allow multiple conforming tools, provide human-readable exports, and test the operating burden across supplier contexts.

A staged test before policy reliance

Government can evaluate graduated verification without granting immediate program effect.

First, freeze representative cases, baseline versions, procedures, measures, success floors, and stop rules. Include technical, procedural, inherited, constrained, degraded, adversarial, changed, and ambiguous cases.

Second, have qualified independent reviewers establish reference findings and document reasonable disagreements. Reference findings are test fixtures; they do not claim universal truth.

Third, assign matched cases to conventional review and proposed routes. Record finding support, missed conflicts, false conclusions, agreement, labor, elapsed time, corrections, and results by reviewer and case type.

Fourth, conduct blind Government re-review of a defined sample. Test whether route assignment and escalation behaved as specified.

Fifth, expand only when a route clears preset quality and safety thresholds. Publish errors and corrections alongside efficiency results. Name permitted uses, required controls, exclusions, and stop conditions.

The pilot cannot prove national-scale performance. It can falsify unsafe assumptions, expose ambiguous rules, and create evidence for the next decision.

Scale should remain reversible

Graduated verification is credible only when Government can observe it and change course. Every route needs versioned rules, audit records, correction and appeal paths, performance measures, and a withdrawal mechanism.

The objective is disciplined allocation of scrutiny. Lower-consequence work can use evidence-backed methods subject to sampling. Intermediate cases can receive qualified assistance and focused human judgment. Highest-consequence work can receive complete independent examination. The Level 2 safeguards remain the same across the design.

CMMC 20X offers this structure for testing. Government should adopt, revise, narrow, or reject it based on measured finding quality, safety, burden, capacity effects, and observed failure modes.

TRACE THE CLAIM

Sources and revision history.

PRIMARY AND GOVERNING SOURCES

01	<u>Cybersecurity Maturity Model Certification Program, 32 CFR part 170</u> <u>Electronic Code of Federal Regulations</u> <u>↗</u>
02	<u>CMMC Assessment Guide — Level 2, Version 2.13</u> <u>Department of War Chief Information Officer</u> <u>↗</u>
03	<u>NIST SP 800-171A Revision 2</u> <u>National Institute of Standards and Technology</u> <u>↗</u>
04	<u>NIST SP 800-30 Revision 1</u> <u>National Institute of Standards and Technology</u> <u>↗</u>
05	<u>CMMC 20X Government Pilot</u> <u>Deep Fathom</u> <u>↗</u>

CORRECTIONS AND MATERIAL REVISIONS

No material corrections recorded.

See an error or a source we missed? [Send a correction](#). Material changes are recorded here rather than silently overwritten.