



ISO/IEC JTC 1/SC 27/WG 3 "Security evaluation, testing and specification"

Convenorship: UNE

Convenor: Bañón Miguel Mr



Assurance_Security_Evaluation_and_Testing_Paradigms_ToR

Document type	Related content	Document date	Expected action
Project / Other	Meeting: VIRTUAL 14 Sep 2026	2026-09-17	

Terms of Reference for a WG 3 Study Period

Title

Study of Assurance, Security Evaluation and Testing Paradigms: evaluation evidence, attack potential, AI-enabled tools, related current and future standards.

Scope

This study period will examine whether the assurance models underlying the WG 3 security evaluation and testing standards remain effective in the light of current reverse-engineering, automated analysis and AI-enabled tools, which are available to evaluators and attackers alike. It uses WG3N3156¹ to set the context, with an expectation that inputs need not be constrained by existing approaches but can include “out of the box” thinking. It may be used to determine which changes to the WG 3 portfolio are justified. The study is intended to be complementary to the PWI WG3N3196 and the two studies will share inputs and be jointly managed.

The study period will:

- a) identify the affected WG 3 standards and the procedural vehicles available for each: preliminary work item, new work item, amendment, technical report or specification, implementation guidance, or next systematic review;
- b) review the evaluation-evidence model of ISO/IEC 15408 and ISO/IEC 18045 and the vendor-evidence model of ISO/IEC 19790 and ISO/IEC 24759. In particular, it will review the premise that information supplied by the developer or vendor gives the evaluator or tester an advantage over the attacker, now that much of that information can be recovered from the delivered product with reverse-engineering and AI tools;
- c) identify inconsistencies between the evaluation assurance level (EAL) scale of ISO/IEC 15408 and the attack potential model of ISO/IEC 18045, such as the EAL scale placing high value on information assets, for example source code, that an attacker can now obtain quickly, cheaply and with little expertise. It will also assess whether the valuation of attack potential factors, and the attacker assumptions underlying ISO/IEC 19790 security levels and non-invasive attack testing, reflect current attack technology, and propose the direction of any refactoring;

¹ISO-IEC JTC 1-SC 27-WG 3_N3156_WG3_AI_Delegated_Evaluation_Contribution

- d) define terminology, roles, autonomy profiles and accountability for AI-enabled execution of security testing and evaluation;
- e) plan, and where possible conduct, controlled pilots in ISO/IEC 15408/18045 and ISO/IEC 19790/24759 contexts;
- f) specify candidate requirements for provenance, reproducibility, containment, qualification and human review of AI-enabled evaluation;
- g) coordinate with CCDB/CCRA, CMVP, scheme bodies that maintain domain-specific attack potential guidance, and existing SC 27–SC 42 liaison mechanisms, including ISO/IEC AWI 26160; and
- h) recommend whether one or more new work items, amendments or implementation deliverables are justified, or whether the study period should be extended on specific topics.

The study will distinguish measured evidence from forecasts. It will not presume the outcome of the review of the current assurance model, and it will not presume that AI output can replace accountable scheme judgement. The evaluation of AI functionality as the target of evaluation is addressed by ISO/IEC AWI 26160 and is out of scope except for coordination.

Call for contributions

National Bodies and liaison organisations are invited to submit contributions, drawing on WG 3 experts, project editors, certification and validation schemes, evaluation and testing laboratories, developers and vendors. Contributions should cite the identifiers of the topics they address. Contributions on related topics within the scope are also welcome.

A. Work programme and procedure

- **A.1 Priorities.** Which topics in this call should become explicit WG 3 work objectives, and in what order.
- **A.2 Procedural vehicles.** Advice on PWI, NWIP, TR/TS, amendment, implementation guidance or next-review pathways for each topic, following publication of the 2026 editions of ISO/IEC 15408 and ISO/IEC 18045 and the 2025 editions of ISO/IEC 19790 and ISO/IEC 24759.

B. Assurance model and evaluation evidence

ISO/IEC 15408 and ISO/IEC 19790 rely on information that the developer or vendor supplies to the evaluator or tester. Examples are the functional specification, TOE design, implementation representation and security architecture evidence of the ISO/IEC 15408 ADV class, and the vendor evidence required by ISO/IEC 24759. One premise of this model is that such information gives the evaluator or tester an advantage over an attacker who holds only the delivered product. Decompilers, firmware extraction, binary lifting, symbolic execution

and fuzzing, especially when combined with AI tools, now let an attacker recover much of that information quickly and cheaply. Contributions are invited on:

- **B.1 Information advantage.** Whether, and for which product types and assurance levels, developer- or vendor-supplied information still gives the evaluator or tester a material advantage over the attacker.
- **B.2 State of the art in reverse engineering.** Evidence on the cost, time, expertise and accuracy of recovering design- and implementation-level information from software, firmware, hardware and cryptographic modules, with and without AI assistance. This includes how far counter-measures such as code obfuscation, encrypted firmware, secure boot and debug-interface locking still raise attacker cost.
- **B.3 Evaluator-derived evidence.** Whether, when and how evidence that the evaluator or tester reconstructs from the product can supplement or replace developer-authored evidence; how such reconstructions should be reviewed; and how to check that the declared design matches the observed implementation.
- **B.4 Value and cost of developer evidence.** Which evidence elements demonstrably help find vulnerabilities or support the assurance argument, and which mainly add cost and duration, including the effect on small and medium-sized developers (see also C.1).
- **B.5 Vendor assertions in ISO/IEC 19790 and ISO/IEC 24759.** Which test requirements rely on vendor documentation or assertions that the tester could now verify independently by analysing the module.
- **B.6 Effectiveness of the model.** Comparison of vulnerabilities found during evaluation or validation with those found after certification or validation, and lessons from certified or validated products later found vulnerable (see also E.2).
- **B.7 Assurance over time.** How fast an assurance result loses value as attack tools improve, and the implications for assurance continuity, flaw remediation (ALC_FLR), surveillance, re-evaluation and patch assurance.
- **B.8 Alternative and complementary models.** For example, source-code access by default, evaluator-led analysis of the delivered product, continuous or incremental assurance, and evidence linked to reproducible builds and software bills of materials.

C. Attack potential model

The attack potential model of ISO/IEC 18045 rates attacks by elapsed time, specialist expertise, knowledge of the TOE, window of opportunity (access to the TOE) and equipment, and maps the result to the AVA_VAN resistance levels. These factors remain broadly sound. The problem lies in how their values are assigned and in their consistency with the EAL scale of ISO/IEC 15408. As assurance rises, the EAL scale requires more developer information, such as the implementation representation (source code), and places high value

on it. Yet an attacker can now often obtain an equivalent representation quickly, cheaply and with little expertise, using reverse-engineering and AI tools. The study needs to identify and revise inconsistencies of this kind. Contributions are invited on:

- **C.1 Consistency between the EAL scale and attack potential.** Where the EAL scale and the attack potential model value the same information or capability differently, for example source code or detailed design that the EAL scale treats as a high-value asset but that an attacker can recover at low cost; and how each inconsistency should be resolved (see also B.4).
- **C.2 Valuation of factor values.** How values within the existing factors should be assigned in the light of current tools. Examples: whether knowledge that can be recovered by reverse engineering should be rated by the effort needed to recover it rather than by its confidentiality classification; how AI assistance affects the expertise and elapsed time an attack needs; and whether general-purpose AI models, cloud compute and low-cost open hardware for side-channel and fault-injection attacks turn specialised or bespoke equipment into standard equipment.
- **C.3 Automated and reusable attacks.** How to rate attacks that, once found, can be automated and reused at almost no cost, and whether a model built around one attacker and one TOE instance should account for automated attacks at scale.
- **C.4 Identification and exploitation.** Whether the distinction between identifying and exploiting an attack, used in scheme guidance, should be kept or revised when AI tools compress or merge the two phases.
- **C.5 AI capability in the model.** How AI-enabled attack capability should be reflected in the existing factors, for example through its effect on expertise, elapsed time and equipment, or through an attacker profile if needed, without depending on specific products or models.
- **C.6 Dated calibration and maintenance.** Whether factor tables should carry a calibration date and a mechanism for periodic recalibration, and what that means for certificates that remain valid over a product's lifetime.
- **C.7 Evaluator-attacker parity.** Whether evaluators and testers should be required to use tools at least as capable as those assumed for the attacker at the claimed resistance level, and how that capability should be demonstrated.
- **C.8 Other WG 3 standards.** Alignment with scheme-level attack potential guidance for smartcards and hardware devices and with ISO/IEC TR 20004, and whether ISO/IEC 19790 security levels, ISO/IEC 17825 and ISO/IEC 20085 show similar inconsistencies between required evidence and attacker assumptions.

D. AI-enabled security testing and evaluation

- **D.1 Terminology and autonomy profiles.** Roles, authority, review and accountability.
- **D.2 ISO/IEC 15408/18045 pilot.** Work-unit mapping and an AVA_VAN case study using controlled TOEs and blinded expert adjudication.
- **D.3 ISO/IEC 19790/24759 pilot.** Machine-readable evidence and automation profiles coordinated with ACMVP.
- **D.4 Qualification testbench.** Reference TOEs, contamination controls, scoring and acceptance criteria.
- **D.5 Continuing proficiency.** Requalification, regression and drift/change monitoring.
- **D.6 Containment requirements.** Network, credentials, scope binding, action approval, logging, emergency stop and incident response.
- **D.7 Evaluator-integrity threat model.** Injection, poisoned evidence, benchmark leakage and harness exploitation.
- **D.8 Evidence and verdict interchange.** A gap analysis that inventories existing formats before any new schema is specified.
- **D.9 SC 42 and ISO/IEC AWI 26160 coordination.** Use of existing liaison mechanisms and a clear boundary between AI as the target of evaluation and AI as an evaluation instrument.

E. Empirical evidence

- **E.1 Evaluation and testing practice.** Anonymised scheme and laboratory data on time, cost, error, reproducibility, review burden and automation pilots.
- **E.2 Vulnerabilities after certification or validation.** Data on vulnerabilities reported in certified or validated products, including how they were found and whether the evaluation evidence could have revealed them.
- **E.3 Attack calibration.** Data on the actual time, cost, expertise and equipment needed for attacks, for example from laboratory attack ratings, penetration tests, bug bounty programmes and AI-enabled vulnerability discovery.

Timeline

The study period runs for six months, from the September 2026 WG 3 meeting to the next WG 3 meeting. Dates will be aligned with the SC 27 meeting calendar and document deadlines. The dates and timeframes below are tentative and will be refined as the study progresses.

Month	Milestone
September 2026 WG 3 meeting	WG 3 approves these terms of reference and appoints the rapporteur(s). The call for contributions is circulated to National Bodies and liaison organisations.
October 2026	Kick-off teleconference. Topic leads assigned. Inventory of affected standards and procedural vehicles (A.2) started.
1 February 2027	Deadline for contributions.
Late February 2027	Teleconference to review contributions. The rapporteur(s) circulate a draft study report.
Mid March 2027	Comments on the draft study report due. Pilot plans and any preliminary pilot results (D.2, D.3) reported.
next WG 3 meeting	Final study report presented, with recommendations on new work items, amendments, guidance, extension of the study period on specific topics, or closure.