Techno Methodology Brief

Technical knowledge and skills assessment by job function

Version 1.3 · 17 November 2026

1. Why this exists

Techno is delivered in two ways:

- **Job-function assessment.** The format most clients receive through an access code: hands-on tasks set in the work of a job function, usually with a short knowledge section, reported in three proficiency bands (section 2).
- **Domain knowledge test.** A multiple-choice test in one of ten technical domains, with a pass mark set by a subject-matter expert for each domain.

This brief explains how both are built, where the items and tasks come from, how results are reported, and how the passing standard is set.

Clients ask three questions about any certification test: *Where did the items come from? How do you know the test covers the job? How was the pass mark set?* This brief answers those questions for the Techno certification program. It follows the structure of the AI Readiness Compass, Assessment Center and Fluent briefs so the product family reads consistently.

In this brief, "certification" means a VIFM-issued technical proficiency credential, valid for one year and verifiable online. It is not a license, a regulated professional qualification or a credential accredited by an external body. The program is designed with reference to the certification standards listed in section 6, but it has not been audited against them. Alignment is VIFM's own mapping against the published standard or law; it is not a certification, accreditation, audit or legal opinion by the body concerned.

2. Construct definitions

Techno measures **job-relevant knowledge and applied skill** within a defined technical domain. The content model is a taxonomy of **domains**, each broken into **skills**. Every item in a domain bank targets exactly one domain and one skill, so a certificate maps to an explicit blueprint of what was tested. (Items written for a job-function assessment are tagged by skill only.)

The construct here is *content-defined*: the test claims to measure mastery of the domain's specified knowledge, not a latent psychological trait. That framing determines which validity evidence matters most - content validity and a defensible standard, rather than factor structure.

The job-function assessment

A job-function assessment is built from the skills of one job function. It has two parts:

- **Hands-on tasks.** The candidate works on realistic material in a spreadsheet, a structured input form or a database query. Each task is scored against weighted checkpoints (for example, a correct value in a given cell, or a query that returns the expected result), and the task score is the weighted share of checkpoints met.
- **Knowledge section.** Where the function has one, a short multiple-choice section drawn from the function's question bank.

The two parts are each scored from 0 to 100 and then blended by a weight set for each job function. The result is reported in three proficiency bands:

Band	Score
Basic	below 60
Intermediate	60 to 84
Advanced	85 and above

A credential is issued only when the knowledge section scores at least 50, the hands-on tasks score at least 50, the blended score is at least 70, the knowledge questions come from the expert-approved bank and every scored task has been approved by a subject-matter expert. A sitting built from a hand-picked subset of a function's skills is indicative only and never issues a credential.

The rest of this brief describes the domain knowledge test unless a section says otherwise.

---

3. Item development

Sourcing and format

Items are multiple-choice questions in four formats: single best answer (four options), scenario-based single best answer (a short case followed by a four-option question), true or false, and select-all-that-apply (four to six options with two or more correct, marked correct only when exactly the right set is selected). Items are optionally bilingual (English and Arabic) and carry an explanation of why the key is correct (a review aid and a source of learner feedback). Each item has a difficulty rating (easy, medium or hard).

Review workflow

Items move through an explicit subject-matter expert (SME) review workflow with five statuses - draft, in review, approved, rejected and retired - and the reviewer's name, the review date and the review notes are recorded. Only approved items are eligible for a live certification test. Each item is recorded as AI-drafted or human-authored; AI-drafted items are never used in a certification test without passing human review. Where a domain or job function does not yet have enough approved items, the platform

can run an indicative test from items that are still in review or that are generated by AI at the time of the sitting; such a result is labeled indicative and does not lead to certification.

Two further controls apply to AI-drafted items. First, numerical items in tests that are generated by AI at the time of a sitting pass an automatic **key verification**: an independent second pass re-solves each item, and any item whose key is not confirmed is dropped (added after reviewers found three wrong keys in a trial). Second, the SME review of the item bank follows a structured procedure: for every item the reviewer records a verdict, their own answer and any correction; approvals and rejections are then applied, while anything that changes a question, an option or a key is held for a person to confirm. Every applied decision is written to a permanent review log that cannot be edited.

Research anchors

Each item is designed to carry a per-item **validation-evidence trail**. When one is recorded, it anchors the item's domain to the content-validity / standard-setting literature (§6) from a curated, closed bibliography, records a confidence level, and is **human-verified** before it reaches any client-facing surface. As of 17 November 2026, none of the 148 approved items yet has a validation-evidence record. Coverage is tracked on VIFM's internal evidence and validity map.

Light psychometrics

Each item accumulates administration statistics (how many times it has been administered and how many times it has been answered correctly), giving a running p-value (difficulty). Discrimination and distractor analysis are planned; they will need each candidate's answer to each item, which the item statistics do not yet hold.

---

4. Validity

Content validity

This is the core defensibility of a knowledge test, and the program's strongest evidence. Every item is domain- and skill-tagged against the taxonomy, and items pass SME review before they can be used. The next documented step is a formal content-validity ratio (Lawshe CVR) panel per domain. The current review trail records one expert's approve, revise or reject decision per item; a CVR study will need several panelists per domain each rating every item as essential or not.

Face validity

Items are reviewed by subject-matter experts for realism and relevance; we mark this *partial* until a structured face-validity review is documented separately from the approval workflow.

Criterion validity and the passing standard

The passing standard is documented per domain: a minimum passing percentage, a defensibility floor on test length (a 3-item "test" cannot certify), and the **method and rationale** for how the standard was set. A domain test is certified only when the domain's standard, method and rationale are all on record; otherwise the test is indicative. A documented, criterion-referenced standard is the appropriate validity

claim for a certification decision; we do not make a separate predictive claim against downstream job performance.

As of 12 October 2026, all ten domains have a standard recommended by one subject-matter expert per domain, who reviewed every item and gave a pass mark with written reasoning based on how a barely competent practitioner would perform (an Angoff-style judgment by a single expert, not a panel). The recorded standards are Accounting, Business Reporting, Finance, Artificial Intelligence and Business Intelligence at 70% and Analytics at 67%, each with a minimum of 8 items, and Banking Operations and Risk Management 78%, Investment 76%, Real Estate 74% and Treasury 73%, each with a minimum of 10 items. Each expert's reasoning is stored with the standard. The experts advised against making any single skill must-pass while a domain has only about 15 items, to be revisited once a domain reaches 30 or more.

Construct validity

Rasch calibration is planned as supporting construct evidence once item banks and response volumes are large enough (item difficulties ordering sensibly on a single scale). It has not yet been run on the certified domain banks. It is treated as supporting, not primary, evidence for a content-defined test.

---

5. Reliability

Internal consistency

Internal consistency (KR-20) has **not yet been estimated** for any domain. It will be calculated and reported per domain once enough candidates have taken the certified test and each candidate's answers to each item are retained for analysis. Until then no reliability coefficient is claimed; the evidence available is the item-development and expert-review process.

Inter-rater reliability

Scoring is **automated** (keyed multiple-choice items), so inter-rater reliability is *not applicable* for scoring. Agreement between judges will become relevant when standards are set by a panel; today each standard rests on one expert, so no agreement figure exists.

Test–retest reliability

Stability across repeated administrations is **not currently tracked** - a disclosed gap. Each attempt draws a new random set of items from the approved bank, which limits memorization of a fixed form but is not evidence of score stability; with banks of about 15 items, two attempts will share several items.

---

6. Reference frameworks and published instruments

Items content-align with the works below; this is content alignment, not republication. Listing a standard here means the item bank was developed with reference to it; it does not mean the program has been

audited against it or is accredited under it. The per-item validation-evidence trail anchors to these.

Content validity and job analysis

- Lawshe, C. H. (1975). *A quantitative approach to content validity.* Personnel Psychology, 28(4), 563–575.
- Raymond, M. R. (2001). *Job analysis and the specification of content for licensure and certification examinations.* Applied Measurement in Education, 14(4), 369–415.
- Kane, M. T. (2013). *Validating the interpretations and uses of test scores.* Journal of Educational Measurement, 50(1), 1–73.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (2014). *Standards for educational and psychological testing.* AERA.
- National Commission for Certifying Agencies (2014). *Standards for the accreditation of certification programs.* Institute for Credentialing Excellence.

Item writing and test construction

- Haladyna, T. M., Downing, S. M., & Rodriguez, M. C. (2002). *A review of multiple-choice item-writing guidelines for classroom assessment.* Applied Measurement in Education, 15(3), 309–333.
- Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory.* Holt, Rinehart & Winston.
- Anderson, L. W., & Krathwohl, D. R. (Eds.) (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives.* Longman.

Standard setting / cut scores

- Cizek, G. J., & Bunch, M. B. (2007). *Standard setting: A guide to establishing and evaluating performance standards on tests.* Sage.
- Angoff, W. H. (1971). *Scales, norms, and equivalent scores.* In R. L. Thorndike (Ed.), Educational measurement (2nd ed., pp. 508–600). American Council on Education.

Published research on the general method

These works are cited in the Reliability and Validity Evidence section. They report published research on work-sample and job-knowledge testing in general; they are not part of the per-item evidence trail and are not studies of Techno.

- Schmidt, F. L., & Hunter, J. E. (1998). *The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings.* Psychological Bulletin, 124(2), 262–274.
- Roth, P. L., Bobko, P., & McFarland, L. A. (2005). *A meta-analysis of work sample test validity: Updating and integrating some classic literature.* Personnel Psychology, 58(4), 1009–1037.
- Roth, P. L., Bobko, P., McFarland, L. A., & Buster, M. (2008). *Work sample tests in personnel selection: A meta-analysis of Black-White differences in overall and exercise scores.* Personnel Psychology, 61(3), 637–661.
- Sackett, P. R., Zhang, C., Berry, C. M., & Lievens, F. (2022). *Revisiting meta-analytic estimates of validity in personnel selection: Addressing systematic overcorrection for restriction of range.* Journal of Applied Psychology, 107(11), 2040–2068.

---

7. Limitations and disclosures

- **Content validity is process-based today.** The SME-review trail supports it; a formal per-domain CVR panel is the next documented step.
- **Cut scores rest on a single expert per domain.** The ten recorded standards were each recommended by one subject-matter expert, as an Angoff-style judgment by a single expert with the reasoning recorded, not a panel. A multi-judge panel per domain is the next step. A domain test is certified only when the domain's standard, method and rationale are on record; otherwise the test is indicative.
- **Item banks are small.** Certified domains hold 14 to 15 approved items each (148 approved items in all, as of 17 November 2026). Tests draw from the whole approved bank, so with banks this size a candidate may see only one or two items on a given skill, which is why no skill is yet mandatory.
- **No verified validation-evidence trail yet.** As of 17 November 2026, none of the 148 approved items has a validation-evidence record.
- **No fairness/DIF analysis yet.** Planned, not run.
- **No test–retest evidence yet.**

---

8. Update cadence

- Item bank: items added, reviewed and retired continuously; item statistics refresh as administrations accrue.
- Cut scores: reset when a domain's blueprint changes; each standard records its method and date.
- Methodology brief: re-issued as the program matures. This document is v1.3.
- Validation-evidence trails: reviewed in the platform's evidence console; coverage is tracked on VIFM's internal evidence and validity map.

---

9. Contact

For methodology questions or research-collaboration inquiries, contact the VIFM consulting team: `contact@viftraining.com`.

Reliability and Validity Evidence

Two kinds of evidence: **(A)** established coefficients for the *method* (work-sample and job-knowledge testing), from the published meta-analytic literature; and **(B)** VIFM's own reliability procedures and their honest current status.

A. Established coefficients for the method (published literature)

The figures below come from published research on these test types in general. They are not results for Techno. VIFM has not collected job-performance or other follow-up outcome data for Techno candidates and makes no predictive claim for it.

- **Work-sample / performance tests:** among the highest-fidelity predictors of job performance. The classic meta-analytic table places work samples at operational validity **ρ ≈ .54** (Schmidt & Hunter, 1998); a more conservative re-estimate reports **ρ ≈ .33** (Roth, Bobko & McFarland, 2005). Both figures are shown; a 2022 re-analysis of the same literature also reports lower values than the 1998 table (Sackett, Zhang, Berry & Lievens, 2022).
- **Job-knowledge tests** (the multiple-choice layer): operational validity **ρ ≈ .48** (Schmidt & Hunter, 1998).
- **Group differences:** earlier reviews suggested that work samples show smaller group differences than cognitive ability tests; a later meta-analysis found larger differences than previously reported (Roth, Bobko, McFarland & Buster, 2008), so no fairness advantage is assumed. Techno's own fairness analysis is planned, not run.

B. VIFM's own reliability - method, threshold, and current status

- **Item statistics:** per-item difficulty (**p-values**) and counts of administrations and correct answers are tracked, behind an SME approve / reject / retire review workflow. Status: light classical statistics only; **IRT / Rasch calibration is the documented next step, not yet run**.
- **Cut scores:** documented per domain with method and rationale; the certified path assembles tests only from SME-approved items and certifies only a score at or above the cut score; otherwise the result is explicitly indicative. Status: all 10 domains have an expert-set standard on record (standards recorded 28 September to 12 October 2026); a domain test is certified only when its standard, method and rationale are on record.
- **Key verification:** numerical items in tests generated by AI at the time of a sitting have their keys re-solved automatically before the test is served; for items in the reviewed bank, the subject-matter expert records their own answer during review; a whole-bank recheck of worked explanations on 28 September 2026 found one wrong key in 148 servable items, since corrected.
- **Delivery integrity:** single-use sessions, server-held answer key, options re-randomized per administration.