Logica Methodology Brief

The cognitive reasoning assessment of the Caliber platform: what it measures, how it is scored, what the report is for, and its scope and limitations.

Version 1.3 · 17 November 2026

1. Why this exists

Whenever a reasoning test informs a people decision, the same three questions arise: *What does the instrument actually measure? How is a single number derived from the answers? And what weight should that number be allowed to carry?* This brief answers those questions for Logica, the cognitive-ability (reasoning) instrument in the Foundations layer of the VIFM measurement model, beneath the observed behavioral assessment of the Assessment Center and the Reflect 360 multi-rater view.

Logica is a foundations-level indicator of reasoning ability. Cognitive ability is a *propensity* - it is theoretically related to how readily a person picks up and applies the behavioral competencies that the rest of the platform measures directly. Logica is deliberately not a hiring decision, not an objective verdict on a person's worth, and at present not a calibrated, norm-referenced test. Every design choice documented below flows from that positioning.

2. What Logica measures

Reasoning, across four subtests

Logica estimates general reasoning ability through four subtests, each a distinct facet of how a person works with information:

- **Numerical reasoning** - working with numbers and quantitative information: ratios and proportion, percentages and percentage change, and interpreting data presented in tables and charts to draw correct conclusions.
- **Verbal reasoning** - understanding and reasoning with language: comprehension of written passages, verbal analogies, and vocabulary in context. Formal logic is measured under deductive reasoning, not here.
- **Inductive reasoning** - inferring the general rule from specific cases: spotting the pattern in number and letter series, figural matrices and odd-one-out problems, then predicting what comes next. A core marker of fluid intelligence.
- **Deductive reasoning** - applying given rules or premises to reach a logically valid conclusion: syllogisms, conditional (if-then) logic, and arrangement or ordering problems where the answer must follow necessarily from the information given.

Items are written to minimize dependence on culturally specific knowledge and to be domain-neutral: they use everyday or abstract content and can be solved without finance, banking or other specialist knowledge. Each item has exactly one defensible correct answer, there are no trick questions, and each item is designed to be solvable in under a minute. A sitting can cover all four subtests or a subset selected when the assessment is set up; if no valid selection is made, all four are administered.

The general (g) composite

When all four subtests are sat, Logica reports a general mental ability (g) composite across them. This is the single broadest signal the instrument produces. Published research on general mental ability tests supports their use as an indicator of how readily people learn; that research concerns the method in general, and VIFM has not yet established this evidence for Logica's own composite. When a sitting covers fewer than four subtests, no g composite is reported (see section 3).

The competencies each subtest relates to

Logica is part of the platform-wide layered measurement model, in which the 21 behavioral competencies form the core and every other instrument relates to that core through a defined type of relationship. Logica's relationship is the one with the weakest evidence. Each subtest is a foundational indicator that is theoretically related to certain behavioral competencies. These relationships are declared on the basis of published research on reasoning ability in general and VIFM's expert judgment; VIFM has not yet established predictive evidence for them in its own data.

- Numerical reasoning is theoretically related to Critical Analysis & Judgment.
- Verbal reasoning is theoretically related to Clear & Adaptive Communication.
- Inductive reasoning is theoretically related to Navigating Complexity & Ambiguity and Creative Problem-Solving.
- Deductive reasoning is theoretically related to Critical Analysis & Judgment.
- The g composite is theoretically related to Critical Analysis & Judgment and Navigating Complexity & Ambiguity.

On the report, these relationships are labeled "Related (foundations)". A strong reasoning score raises the expectation that a person will pick up the related competencies readily; it is not evidence that they currently demonstrate them. Only the observed instruments (the Assessment Center, competency-based interviews and Reflect 360) *manifest* a competency.

---

3. How it is scored

Format and timing

A sitting drawn from the item bank has nine items per subtest: one easy, one medium and one hard item from each of the subtest's three content areas, so 36 items for the full battery. The default time limit is 40 minutes for the full battery; for a shorter standalone sitting it is reduced in proportion to the number of subtests, with a minimum of 10 minutes. The limit can be changed by the administrator.

From items to a subtest score

Every Logica item is a multiple-choice question keyed to exactly one subtest, with one correct option. A subtest's raw score is the percentage of its items answered correctly: the number of correct answers divided by the number of items in the subtest, multiplied by 100 and rounded to a whole number. An unanswered item counts as incorrect, no marks are deducted for a wrong answer, and the speed of response does not enter the score. When a sitting is timed, it is submitted automatically at the time limit, and any item still unanswered counts as incorrect. The same percentage serves as the subtest's 0-100 score.

From subtests to the composite

The composite is the mean of the percentages of the subtests in the sitting, rounded to a whole number. It is reported as a 0-100 composite, not as an IQ-style standard score. How it is presented, in the result report and on every screen that shows it, depends on how many subtests were sat:

- **All four subtests** - the composite is reported as the general mental ability (g) composite.
- **Two or three subtests** - the composite is the mean of the subtests sat and is labeled "Reasoning average (partial battery)". It is not presented as g.
- **One subtest** - no composite is shown; the report gives the subtest result only.

Interpretation bands (indicative)

Each subtest percentage, and the composite, is interpreted against five indicative bands based on raw percentage correct, not on a peer distribution:

- **Very high** - 90 and above
- **High** - 75 to 89
- **Moderate** - 60 to 74
- **Low** - 40 to 59
- **Very low** - below 40

The bands are fixed cut-offs on the share of correct answers, not a comparison with other people. They are indicative and do not compare the taker with a population or a norm group.

Each subtest also carries a short, non-clinical narrative that matches its band, so the band and the narrative always agree and a reader sees a plain-language description alongside the number.

Norm-referencing - provisional, and gated

Logica is engineered to become norm-referenced - to report a percentile and a 1-10 sten score relative to a VIFM reference group - but only once the evidence to do so exists. The gate is explicit. An instrument is promoted from Indicative (Tier 1) to Calibrated (Tier 2) only when, for every scale, three thresholds are cleared at once:

- at least 9 approved items per subtest (content): one full form, with one item for each content area and difficulty level,
- internal-consistency reliability (Cronbach's alpha) of at least 0.70, counted toward the gate only when it is computed on at least 200 complete responses (the same as the norm sample), and
- a comparison (norm) sample of at least 200.

The scoring step applies all three at once, for every scale: a percentile is reported for a scale only when its comparison sample is at least 200, its alpha is at least 0.70 on at least 200 complete responses and at least 9 of its items are approved. Until a scale clears all three, its norm is *ignored* in scoring even if pilot data is accumulating, so percentiles are not reported. As of 17 November 2026, no reasoning item yet carries the approved status and no norm sample exists, so no scale meets the content threshold (on 16 November 2026, 124 items were in review, 6 were drafts and 21 were retired). The figure of 200 is a minimum gate, not a guarantee that a norm group is adequate: the group must also be representative of the people being assessed, and validated norms require review and sign-off by a qualified psychometrician. When the gate is met, a raw score is standardized against the norm group's mean and standard deviation into a z-score, a percentile (via the standard normal distribution), and a sten band; a result is marked "calibrated" only when every scale in the sitting, and the composite, has a norm sample that meets the minimum. With no norm group that meets the minimum, the result is left exactly as a Tier 1 indicative band. **At the time of writing, Logica runs at Tier 1: indicative bands only, no percentiles, no sten, no credential.**

---

4. Deliverables

The professional result report

A completed Logica sitting produces a professional PDF result report (English). It opens with a clear tier badge - "Tier 1 - Indicative" or "Tier 2 - Norm-referenced" - so the report cannot overstate its own rigor, and a "How to read this report" panel that states plainly whether the scores are raw-score bands or norm-referenced percentiles.

Each subtest is rendered as a card showing the raw percentage and band, the fuller definition of what the subtest measures, the score-band narrative, and the competencies that subtest is a foundational indicator for (labeled "Related (foundations)"). The composite is shown as its own card, labeled as described in section 3; a single-subtest sitting has no composite card. When, and only when, the result is calibrated, each subtest card adds a percentile, a sten band (1-10) and a percentile bar, and the composite card adds a percentile; in the indicative state those are deliberately absent.

A second page explains how to read the result responsibly: a "What this is - and isn't" panel stating that reasoning ability is a foundational indicator for behavioral competency and does not measure it directly, that the result is one input to a human judgment and never an automatic decision, and that it yields a score and a profile, not a pass/fail credential. A "Methodology and limits" panel restates how the score was computed and what validated high-stakes use would additionally require.

Where the score flows

A Logica result can be bound to a candidate and an engagement, in which case its competency-related signals appear on that candidate's skills profile as foundations-layer signals, marked as related to the competency and not a measure of it. The result feeds human judgment in development planning, selection short-listing, and succession discussion. It is never used downstream as an automatic gate. Where an engagement combines Logica with other instruments, for example the Emergia high-potential profile, the reasoning score contributes to that profile's ability measure and placement according to that service's own published method.

---

5. What Logica is, and what it is not

- **It is a foundations-level indicator of reasoning ability.** It estimates how readily a person reasons with numbers, language, patterns and rules. That is a useful foundational signal - but it is one input.
- **It is a foundational indicator for competencies; it does not measure them.** Logica sits at the propensity tier, the weakest-evidence relationship in the layered model. A high reasoning score raises the expectation that related competencies will develop readily; it is not evidence that they are currently demonstrated.
- **It is Tier 1, indicative.** There is no local norm group, so scores are raw-score bands, not percentiles. There is no IRT or Rasch calibration of the item bank. The same instrument becomes norm-referenced (Tier 2) once a representative sample and reliability evidence accumulate. Results already issued remain indicative unless they are re-scored against the norm group; takers do not need to be re-tested.
- **It is not a credential.** Logica yields a score and a defensible report. It does not issue a pass/fail certificate, and it has no reject path.
- **It is not a hiring decision.** A reasoning score informs a human conversation; a human makes every decision. Validated high-stakes selection use additionally requires a criterion-validity study, adverse-impact (fairness) analysis, and a qualified psychometrician's sign-off.

These caveats are not an afterthought: the tier badge, the statement that reasoning indicates, and does not measure, competency, and the "score, not a credential" framing appear on every report.

---

6. Integrity and secure delivery

Because the score is only as trustworthy as the conditions it was produced under, Logica uses the same secure-delivery model as the platform's other objectively scored instruments (Fluent, Techno).

- **The answer key never reaches the browser.** When a sitting starts, the full test, including the correct answer for each item, is assembled and held on the server. The taker's device receives only the questions and the answer options, so the correct answers are never present in anything the device can read.
- **Grading happens on the server.** On submission, the server scores the submitted answers against the test it holds and computes the bands and the composite. The browser is never trusted to grade itself.
- **Sessions are single-use.** A session is closed the moment it is scored; a second submission against the same session is refused. A scored session cannot be scored again, which prevents a session from being replayed to obtain a better outcome.
- **Further controls.** Answer options are reordered for each sitting; question order is randomized within each subtest; items are rotated so that the least-used items are served first; an invitation can be completed once; a server-side time limit backs up the on-screen timer; copying is disabled during the test; and the report carries a "Preliminary results" notice while any subtest does not yet have enough approved items.
- **Results are read-restricted.** Results can be read only by authorized VIFM staff and, for results that belong to their own organization, by an authorized client manager. Results are written only by the server, never by the taker's device. An invited taker receives a confirmation that the assessment was submitted, not a score or a report; only an authorized user can download the report. Item-level responses are logged separately as the basis for future calibration.

Two limits accompany this model. First, secure delivery is not proctoring: the integrity controls protect the answer key and prevent replay, but they do not by themselves confirm the identity of the taker or that the person worked unaided. Second, the items are AI-drafted. A sitting bound to a candidate, an engagement or an invitation is assembled only from Logica's item bank, which holds items approved in expert review together with items still awaiting review; a demonstration run that is not bound to a candidate may instead use items generated at the time of the sitting or a short built-in set. Until a scale's bank is sufficiently populated, reviewed and calibrated, Logica stays in its indicative state.

---

7. Fairness, proper use, confidentiality and retention

Standards the instrument is held to

Logica's design and intended use are aligned with the following recognized standards for the proper use of assessment:

- **ITC Guidelines on Test Use** - on the responsibilities of those who select, administer and interpret assessments.
- **Standards for Educational and Psychological Testing** (AERA / APA / NCME, 2014) - referenced for evidence-based item development and for the discipline of matching claims to the evidence, including the discipline of not reporting norm-referenced scores until a defensible norm group exists.
- **Uniform Guidelines on Employee Selection Procedures (1978)** - used as a reference for how a cognitive instrument may and may not be used in a selection context. Where Logica is used for selection, the user organization remains responsible for monitoring adverse impact and for the validity evidence the Guidelines call for. Cognitive-ability tests warrant particular care for adverse impact; Logica's positioning as one input to a human decision, with no reject path, is deliberate.

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.

Proper use

The propensity framing is applied consistently across the instrument. Logica indicates, it does not measure; it informs a human judgment, it does not decide; it produces a score, not a credential. The fairness logic is structural rather than cosmetic: there is no automated cut-off in the scoring, the report itself states that reasoning indicates, and does not measure, competency, and norm-referenced scores are gated behind a minimum norm sample, with content and reliability thresholds set for promotion to Tier 2, so that the instrument does not overstate its evidence. Validated high-stakes use additionally requires a criterion-validity study, an adverse-impact analysis, and a psychometrician's sign-off - none of which Tier 1 asserts.

Confidentiality

A Logica result is confidential personal data. Results can be read only by the authorized users described in section 6; a taker without an account cannot retrieve a report by guessing an identifier. Where a result is bound to a candidate and engagement, it is handled within the same access controls as that candidate's other assessment data.

Consent

Consent is recorded before any response is collected, on every delivery route. Standalone access codes issued before 17 November 2026 are completed under the terms they were issued with.

Retention

Logica data follows the platform's default data-retention rule - a maximum of two years unless a contract extends it - designed with reference to UAE Federal Decree-Law No. 45 of 2021, the Saudi PDPL, and GDPR where applicable.

8. Bilingual availability

Logica is delivered bilingually. Items, options, subtest definitions, score-band narratives and interpretation bands exist in English and in Modern Standard Arabic, and a taker sits the test in either language on one consistent scoring basis - the same subtests, the same answer keys, the same composite. Items in the bank are held in parallel English and Arabic versions, and an Arabic sitting is assembled only from items that have a complete Arabic version; items generated at the time of a demonstration sitting are written directly in the requested language rather than translated. The professional result report PDF is currently English-only; a bilingual report can follow when it is prioritized. Arabic content is best-effort pending human review before high-stakes use.

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* (general mental ability testing), from the published meta-analytic literature; and **(B)** VIFM's own reliability procedures and their current status.

A. Established coefficients for the method (published literature)

- **General mental ability (GMA):** among the most widely studied predictors of job performance. Operational validity **ρ ≈ .51** in the classic meta-analysis (Schmidt & Hunter, 1998); a 2022 reappraisal, which found that earlier corrections for range restriction had overstated validity, re-estimates GMA validity at **ρ ≈ .31** (Sackett, Zhang, Berry & Lievens, 2022). Both are cited: the range across the two is roughly **.31-.51**.
- **Internal consistency:** published full-length reasoning batteries commonly report good internal consistency. Reliability depends on test length, so a 9-item Logica subtest should be expected to be lower until it is measured.

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

- **Tier:** Logica is currently **indicative** - AI-generated items, no local norms, no IRT calibration yet, and it issues no credential.
- **Internal consistency:** Cronbach's **α** per subtest, target **α ≥ .70**, counted only when computed on at least 200 complete responses; **Rasch / IRT difficulty calibration** is the documented Tier-2 path. Status: not yet met; as of 17 November 2026 no item is approved and no norm sample exists.
- **Norms:** a norm sample of at least **n = 200**, together with alpha of at least .70 (computed on at least 200 complete responses) and at least 9 approved items for the scale, is the minimum gate before raw scores are converted to percentiles and stens; it is a threshold, not a guarantee that the norms are adequate. Until then only indicative bands are shown.

References

- 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.
- 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.