GE-TECH-1.0 / VERSION 1.0

Technical Review & Calibration Plan

Organizational performance and service quality

Great Entity Institute
2778 Olaya St.
Riyadh, Saudi Arabia

Document date: 26 October 2026
Assessment and certification standards.

In this document

01 Technical desk review

02 Findings and design risks

03 Field pilot and calibration protocol

04 Calibration case set

05 Study record and responsibilities

Use this document with its stated edition. Applications and enquiries: contact@greatentity.com. Certificate verification: verify@greatentity.com. Certification follows an assessment and an authorized decision.

Technical desk review

Review of the authored GEEF 1.0 design and its operational implementation.

What this review establishes

This review checks internal consistency, the connection between criteria and decisions, the treatment of evidence, and the implementation of decision controls. It is an analytical review of an authored scheme. It does not establish predictive validity, independent expert endorsement, or performance in completed field assessments.

Design reviewed

Twenty criteria, grouped into five equally weighted dimensions. Each criterion is worth 50 points. Approach, implementation, results, and learning are rated 0 to 4 with weights of 20%, 30%, 30%, and 20%. Certification requires 800 / 1,000 overall, at least 120 / 200 per dimension, the mandatory gates, and no open critical or major finding.

Operational controls

Each application identifies scope, locations, sector, and specialist programs. The applicant acknowledges scope. Ratings cite uploaded evidence. Submitted assessments are reviewed by a different assigned person. Approval creates the certificate and register entry. Role restrictions and a decision audit trail support accountability; they do not prove assessor competence by themselves.

Findings and design risks

Issues to test before making claims about validation or cross-sector comparability.

Threshold and weighting

The 800-point threshold, 120-point dimension floor, and component weights are scheme design decisions. Arithmetic consistency does not show that these cut-offs distinguish excellence in practice. Test score sensitivity around the threshold and compare borderline cases with independent judgments of the evidence.

Ordinal ratings

Levels 0 to 4 describe ordered performance categories. Treating adjacent levels as equal numerical steps is a practical scoring convention, not a measured interval scale. Avoid claims of precision such as one organization being exactly a stated percentage better than another.

Potential double counting

Results and learning recur across dimensions. Require assessors to identify the distinct claim supported by each piece of evidence. The same document may support multiple criteria, but the report must not reward the same result repeatedly without a different criterion-specific reason.

Feedback sufficiency

A baseline count of 20 responses cannot by itself establish representative feedback. Review population size, selection method, response rate, channel coverage, exclusions, and nonresponse. For small populations, document coverage and use triangulation rather than inventing additional respondents. Escalate any departure from the published requirement for an explicit scheme decision.

Equal dimensions across sectors

Equal weights make the framework easy to explain but may hide material sector risks. Mandatory modules and critical-finding gates must prevent a high average score from offsetting a serious failure. Test this with adverse cases in each sector.

Paperwork bias

A polished policy is not evidence that a process works. Include cases with excellent documents but weak delivery, and cases with modest documentation but traceable, effective practice. Assessors must distinguish a documented approach from implementation and results.

Claims supported today

The institute can explain its criteria, calculation method, evidence requirements, and decision process. Claims of proven reliability, benchmark superiority, international recognition, or validated thresholds require separate evidence. Record that evidence before using such claims.

Field pilot and calibration protocol

A proposed study to generate evidence; no pilot outcomes are claimed.

Recruitment and coverage

Begin with six consenting organizations across at least three sectors and two size bands. Include one small organization, one multi-site operation, and one digitally delivered service. This is a practical learning sample, not a statistically representative validation study. Obtain written agreement on data handling, report use, and whether the pilot has certification consequences.

Independent rating

Use two trained assessors per case. Both receive the same evidence pack and scope but rate independently before discussing results. Retain original ratings, reasons, uncertainties, and time spent. A third competent reviewer moderates differences after the independent records are locked.

Measures to record

For every criterion, record exact agreement, agreement within one rating level, the direction and size of disagreements, missing evidence, and interpretation questions. Report dimension and total score differences, and whether the assessors reached different certification outcomes. Where the sample supports it, calculate an appropriate agreement statistic with uncertainty and explain its limitations.

Working acceptance signals

Proposed internal triggers: investigate any criterion with a two-level difference; investigate total-score differences above 50 points; and resolve every disagreement that changes certification eligibility. These are review triggers, not externally validated reliability standards. Do not tune them after seeing results merely to produce a pass.

Sensitivity checks

Recalculate borderline cases with modest alternative component weights and with each contested rating moved by one level. Identify decisions that change. Examine whether one sector, organization size, channel, or evidence style systematically receives different treatment without a defensible reason.

Release decision

The technical lead compiles findings, proposed revisions, and unresolved risks. An independent competent review group examines the evidence and records the release decision. Revise anchors and guidance where needed, repeat affected calibration cases, and publish the edition and transition arrangements. Actual reviewers and approvals must be recorded; this plan does not appoint them.

Calibration case set

Use common evidence packs to test reasoning, not memorized scores.

Case A - policy without implementation

Supply a complete policy pack alongside transaction samples showing inconsistent execution. Ask assessors to distinguish approach from implementation and results. Expected discussion: documents cannot substitute for observed practice.

Case B - small population

Supply a service with 12 beneficiaries and feedback from ten, with documented attempts to reach all twelve. Compare with a service that has 10,000 beneficiaries and 20 convenience responses. Expected discussion: respondent count alone does not establish coverage or representativeness; scheme departures need explicit handling.

Case C - high total, serious failure

Supply strong scores across most criteria but evidence of a material unresolved control failure. Expected decision: a critical or major finding blocks approval even where the numerical threshold is met.

Case D - borderline score

Supply a record near 800 with one criterion supported by contradictory sources. Require independent ratings and reasons before moderation. Expected discussion: avoid adjusting a rating to reach a desired decision.

Case E - corrective action

Supply a revised procedure, training attendance, and a later sample that still shows the original failure. Expected decision: action completion is not verified effectiveness; the finding remains open.

Case F - scope and exclusions

Supply a strong main site and an excluded branch that delivers the same service under the same proposed certificate. Expected discussion: assess the legitimacy of the exclusion, prevent misleading scope, and record the boundary explicitly.

Study record and responsibilities

Complete these records with actual people and evidence when the study is run.

Required case record

Case identifier; consent; sector and size; scope; evidence pack version; assessor competence records; conflict declarations; independent ratings; rationale; time spent; moderation record; proposed changes; final decision; and limitations.

Required scheme record

Document owner; technical lead; independent reviewers; study dates; case selection rationale; data retention agreement; analysis file; interpretation decisions; change log; approval record; publication date; and transition instructions.

Separation of responsibilities

The technical lead administers the study and consolidates results. Assessors rate cases independently. Reviewers challenge the interpretation and proposed changes. Commercial staff must not override technical findings to improve a pass rate.

Completion standard

The plan is complete when the case records, analyses, reviews, and decisions exist and can be inspected. A written protocol, a software test, or a sample report cannot substitute for running the study. Until then, describe the work as a planned field evaluation rather than a completed validation.