

Designing Criterion-Referenced Assessment

Rubric with anchor exemplars · per Jonsson & Svingby (2007) on exemplar-driven assessor calibration, criterion-referenced standards per Sadler (1989).

Criterion	Developing	Proficient (target)	Distinction
Criterion writing (Bloom's)	Trainer writes: "Learner will understand fillet welds." Verb is non-observable; no product specified; no condition or tolerance. Two assessors cannot agree on what evidence would satisfy this criterion.	Trainer writes: "Learner will produce a 6mm horizontal fillet weld on mild steel to BS EN ISO 5817 Level C, within 8 minutes, without external prompting." Verb observable, product specified, tolerance bounded.	Trainer writes the Proficient criterion AND adds a discriminating Bloom's "Evaluate" step: "...then justify, in 60 seconds to the assessor, why their travel angle compensated for the joint's heat sink, citing one alternative they rejected."
SOLO taxonomy grading	Learner asked why their RCD tripped during termination test answers: "Because it went off." Single fact recalled in isolation. Trainer grades this as Pass because "they got the right answer" — SOLO unistructural mis-graded as competent.	Same learner answers: "The RCD tripped because I had a neutral-earth fault on the load side — I'd nicked the neutral sheath when stripping." Two facts linked causally. Trainer grades as SOLO multistructural, correctly.	Learner answers: "Neutral-earth fault from a strip nick. I'd also check the upstream RCD selectivity because this circuit feeds a freezer — nuisance tripping there has a cost." Causal chain + transferred principle = relational/extended abstract.
Threshold standard preservation	Two learners present identical refrigerant gauge-bleeding procedures with a 4-bar pressure overshoot. Trainer passes Learner A ("good effort, nervous on the day") and refers Learner B. Threshold has drifted with learner narrative, not evidence.	Trainer references the published threshold ("overshoot >2 bar = refer") and refers both learners identically, logging the bar reading in each portfolio. Threshold holds across the cohort on the same day.	Trainer holds the threshold across the cohort AND across time: produces last year's portfolio sample at the 2-bar boundary as an anchor reference, demonstrating year-on-year drift control to the IQA / examencommissie.
Workshop transfer	Hairdressing assessment: learner sections a mannequin head into four perfect quadrants. Trainer awards full marks. No evidence sought that the learner can re-section on a live client whose crown is asymmetrical.	Same task, but assessment requires the learner to section one mannequin AND one live model with a documented cowlick, applying the same sectioning logic. Transfer to one varied context is evidenced.	Learner sections the live model, then verbalises to the assessor how they would adjust the section for a client with traction alopecia at the temples — transfer to an unseen context the learner has named themselves.

Anchor exemplars exemplify what each level actually looks like in evidence — not what it should be. Two assessors using these anchors independently should reach the same grade on the same evidence (inter-rater convergence target).

Regional standards bridges. The RCD anchor cites BS 7671 (IET, UK). Equivalent reference for NL trainers: NEN 1010 art. 41.3 (aardlekschakelaars). Equivalent for DE trainers: DIN VDE 0100-410 §411. The pedagogical logic (unistructural → multistructural → relational) transfers across all three; the citation differs. A bilingual rubric pack is on the roadmap once the NL/DE validator pool is named.

Access & fairness note (Workshop Transfer row). The Distinction asks the learner to name an unseen context (e.g. traction alopecia at the temples). This is a deliberate transfer prompt — but the cognitive load shifts to the learner. For candidates with reasonable adjustments under JCQ access arrangements (2026 ed.), the trainer should offer the unseen context as a prompt-card option rather than a self-generated recall task, preserving the same evidential threshold without disadvantaging the learner.

Assessor Calibration Exercise

To be run between two trainers immediately after Module 3. Bloxham, den-Outer, Hudson & Price (2016) framing on socially-constructed assessor standards; aligned with OfQual moderation guidance on inter-rater convergence.

Purpose

To surface where two trainers diverge in applying the same rubric to the same learner evidence, and to agree one calibration action for the next cohort. This is not a quality-assurance audit; it is a standards-construction conversation.

Materials to bring

- One piece of authentic learner evidence each (photographed weld, portfolio termination sheet, recorded client-consultation transcript, refrigerant log).
- The Module 3 rubric (4 criteria × 3 levels, this sheet).
- A blank grading sheet per trainer.
- A pen each. Phones face-down.

Step 1 - Independent grading (5 min)

Swap evidence pieces. Grade your partner's evidence against all four rubric criteria in silence. Record one level per criterion (Developing / Proficient / Distinction) and write a one-line justification per criterion citing the specific feature of the evidence that drove your decision (e.g. "3mm undercut on the toe = Developing on Threshold").

No talking. No glancing across.

Step 2 - Comparison and disagreement surfacing (5 min)

Reveal grades simultaneously. For every criterion where you diverged by one or more levels:

- Each trainer reads aloud their one-line justification.
- Ask: "*What feature of the evidence did you weight that I did not?*"
- Ask: "*Is our disagreement about the evidence, or about the rubric wording?*"

Do not negotiate to a middle grade. Disagreements are data, not problems. Write the divergence verbatim in the margin of your rubric.

Step 3 - Convergence on next-cohort calibration (5 min)

Together, name **one** of the following as your calibration commitment for the next cohort you teach:

1. **Rubric edit** — the wording at one level is ambiguous; agree a sharper anchor exemplar to add.
2. **Evidence-collection edit** — the evidence type is under-specified; agree what learners must additionally produce so the rubric can be applied.
3. **Threshold re-anchoring** — both of you were lenient or strict on the same criterion; agree the boundary case you will hold next time.

Write the commitment on both rubric sheets. Date it. Bring it to the next moderation cycle.

Closing reflection prompt (individual, ~3 min in your CPD log)

"On which criterion did I most trust my professional judgement over the rubric wording — and was that trust warranted by the evidence, or was it habit?"