

Validity, Reliability, Fairness and Inclusivity in Assessment

Category: Assessment, Feedback and Learning

Validity, reliability, fairness and inclusivity are interdependent qualities of defensible assessment. Validity concerns whether evidence supports the intended interpretation and use of a result. Reliability or precision concerns the dependability of evidence and judgement. Fairness and inclusivity concern equitable opportunity to demonstrate the intended learning without irrelevant barriers.

These qualities cannot be secured by a checklist or by uniform treatment alone. Standardisation may support reliability, yet an inaccessible format can introduce construct-irrelevant difficulty and weaken validity. Flexible conditions or modes are defensible when they preserve what the assessment is meant to assess ([AERA et al., 2014](#)).

Four Interdependent Qualities

1. **Validity:** Define the construct, decision and evidential argument, then investigate alternative explanations.
2. **Reliability:** Use sufficient sampling, clear procedures, assessor preparation and moderation to reduce avoidable inconsistency.
3. **Fairness:** Examine bias, treatment, access, consequences and routes for challenge or correction.
4. **Inclusivity:** Anticipate learner diversity and remove unnecessary linguistic, cultural, sensory, technological or procedural barriers.
5. **Integrated review:** Evaluate outcomes across relevant groups and revise tasks, conditions, criteria and uses together.

Examples in Practice

A clinical-reasoning assessment samples several cases, uses trained assessors and moderated criteria, provides accessible materials and approved adjustments, and tests whether scores differ for reasons unrelated to clinical reasoning. The standard remains constant while irrelevant obstacles are reduced.

Why It Matters

Reliable scores can support an invalid conclusion, and formally identical conditions can be inequitable. Assessment quality therefore requires an evidence-based argument about meaning, consistency, access and consequence. Inclusive design strengthens rather than dilutes validity when it enables diverse learners to demonstrate the intended construct ([Tai et al., 2024](#)).

Further Reading

6. AERA, APA, & NCME. (2014). [Standards for educational and psychological testing](#). American Educational Research Association.
7. Messick, S. (1995). [Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning](#). American Psychologist, 50(9), 741–749.
8. Tai, J., Ajjaw, R., & Umarova, A. (2024). [How do students experience inclusive assessment? A critical review of contemporary literature](#). International Journal of Inclusive Education, 28(9), 1936–1953.

Related Terms: Assessment Validity | Assessment Reliability | Inclusive Assessment | Reasonable Adjustments