Psychometric tests

Lesson 7.2.4 | A Level issues, debates and methods

Core ideas

A psychometric test measures a psychological characteristic using standardised tasks, questions or scoring rules.

Standardisation means the test is given and scored in the same way so that scores can be compared fairly.

Norms are comparison data from a relevant group, helping researchers interpret whether a score is typical, high or low.

Reliability matters because a test should give consistent scores when the construct has not changed.

Validity matters because the test should measure what it claims to measure, not something unrelated such as reading speed or test anxiety.

Psychometric tests can support decisions in education, work and clinical settings, but misuse can label people unfairly or ignore culture and context.

Key terms

Psychometric test: A standardised test designed to measure a psychological characteristic.

Standardisation: Using the same instructions, materials, timing and scoring for each test taker.

Norms: Comparison scores from a relevant group used to interpret an individual's result.

Psychometric tests

Lesson 7.2.4 | A Level issues, debates and methods

Continued

Reliability: The consistency of a test or measure.

Validity: How far a test measures what it claims to measure.

Raw score: The score before it is compared with norms or converted into another scale.

Percentile: A score showing the percentage of the norm group scoring at or below that point.

Test bias: A feature of a test that unfairly disadvantages some groups.

Further detail and evidence

Psychometric tests systematically measure psychological attributes using defined items, administration and scoring. Standardisation supports comparable use. Norms, where relevant, provide a reference distribution; a criterion threshold instead compares with a defined standard.

Reliability includes internal consistency, test-retest and rater agreement where relevant. Validity includes whether items represent the intended construct and relate to suitable external evidence. High consistency can arise from repetitious items without covering the whole construct.

Examples include BDI depressive symptom scores, GAD-7 anxiety severity and JDI satisfaction facets. Some clinical ratings combine interview and judgement; not every named scale is a purely self-administered test. A score must be interpreted using the correct version and intended population.

Psychometric tests

Lesson 7.2.4 | A Level issues, debates and methods

Applying the ideas

A total can monitor change before/after an intervention, but consider whether the instrument is sensitive to the relevant symptoms. A high screening score supports further assessment; it does not automatically establish diagnosis, cause or the best treatment.

Worked example: A school uses a standardised aptitude test to decide who receives extra enrichment lessons. What should be checked before relying on the scores? The school should check that the test is reliable, valid for the ability it claims to measure and standardised for the age group being tested. Norms should come from a relevant comparison group, not a very different population. The score can help decision-making, but it should not be the only evidence because motivation, language background, anxiety and classroom performance may also matter.

Worked example: When a personality test is translated for students in another country, several items about workplace habits no longer make sense. What problem does this create? This creates a validity and test-bias problem because the translated items may not measure the intended personality trait in that cultural context. Students may answer based on unfamiliar wording or irrelevant examples rather than their actual characteristics. The test would need careful adaptation, piloting and comparison with suitable norms before the scores were interpreted seriously.

Evaluation and connections

Tests are efficient and can reduce unsystematic judgement, while response bias, language, culture and inappropriate norms threaten interpretation. Repeated measurement can introduce memory or practice. Examine practical access and confidentiality. A reliable instrument used for the wrong purpose can still lead to an invalid conclusion.