

Reliability

Education

Whether the same pupil would get the same result on another day, paper or marker.

Reliability is the consistency of a measurement. Every score is the true score plus error, and the error is larger than most reporting admits.

REPORTED AS

0 to 1

COMMON TARGET

0.8 and above

- Longer tests are more reliable than short ones.
- Reliability sets a ceiling on validity.
- Test-retest, internal and inter-rater forms differ.
- Single questions are very unreliable.
- Small score differences are usually noise.
- A reliable test can still measure the wrong thing.

LOOK FOR IT

A pupil scoring 62 on a test with a standard error of 4 marks could plausibly have scored anywhere from about 54 to 70.

Educational reference. Grade boundaries fall inside the error band, so a proportion of grades near any boundary are effectively arbitrary.