

Webinar II: How Low Reliability Affects Scoring

Reliability of measurement is extremely important. It is integral to validity. Conceptually, you can't draw valid conclusions about the results of a survey or test if the data aren't reliable. And technically, in the calculation of validity coefficients, the degree of reliability in a set of scores puts a limit on the ceiling for their validity. For that reason, the National Standards for Educational and Psychological Tests does not treat reliability as something separate from validity, but as an intrinsic part of the validity argument.

One way to appreciate the importance of reliability is illustrated in the table below. It gives you an idea of how much scores are likely to change with repeated testing, depending on the reliability of the instrument.

**Percent of Scores that Would Change upon Re-testing
With Different Levels of Test Reliability¹**

Test Reliability	Change 0.5 SD	Change 1.0 SD	Change 1.5 SD
r = 0.50	62%	32%	13%
r = 0.70	52%	20%	5%
r = 0.80	43%	1%	2%
r = 0.90	26%	3%	1%

SD = standard deviation, a measure of the amount of variance in your data

In the left-hand column you see ranges of test reliability, from 0.50 to 0.90. The column headers to the right show three levels of estimated change, as measured in standard deviation (SD) units. The numbers in the cells represent the expected percentage of scores that would change by at least half a standard deviation, or a full standard deviation, or 1½ standard deviation units. Let's read across the row data, starting with a test with only r = 0.50 reliability. The table tells us that for a test with 0.50 reliability, 62% of the people taking the same test again would find they got a different score – and that score would be half a standard deviation different than their first score. A third (32%) of the test-takers' scores would vary by a full standard deviation the second time around, and 13% by 1½ standard deviations. Compare these findings with those for tests with r = 0.90 reliability.

The bottom line: changes in scores on retesting can be quite large, especially for tests with lower levels of reliability.

¹Wainer & Thissen, as shown in Downing, S.M., *Medical Education*, 2004: 38:1006-12.