

What is Test-Retest Reliability?

Authored by
stats writer

April 26, 2024

RECOMMENDED CITATION

stats writer (2024). *What is Test-Retest Reliability?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=139663>

Test-Retest Reliability refers to the consistency or stability of a measurement or test over time. It is a statistical method used to determine if the results of a test are consistent when administered multiple times to the same group or individual. This reliability is important as it ensures that the test or measurement is accurate and not influenced by external factors. For example, if a psychological test is given to a group of individuals and then repeated after a certain period of time, the results should be similar if the test has high test-retest reliability. This indicates that the test is reliable and can be used to make consistent and accurate assessments.

What is Test-Retest Reliability? (Definition & Example)

Often researchers want to use some type of test to measure a construct like intelligence, aptitude, educational ability, etc. in individuals of some population.

When administering any type of test, it's important that the test has reliability. In other words, it's important that the results of a test can be reproduced under the same conditions at two different points in time.

Test-retest reliability is a specific way to measure reliability of a test and it refers to the extent that a test produces similar results over time.

We calculate the test-retest reliability by using the , which takes on a value between -1 and 1 where:

-1 indicates a perfectly negative linear correlation between two scores

0 indicates no linear correlation between two scores

1 indicates a perfectly positive linear correlation between two scores

For example, we may give an IQ test to 50 participants on January 1st and then give the same type of IQ test of similar difficulty to the same group of 50 participants one month later.

We could calculate the correlation of scores between the two tests to determine if the test has good test-retest reliability.

Test-Retest Reliability = Correlation between test scores



Generally a test-retest reliability correlation of at least **0.80** or higher indicates good reliability.

Example: Calculating Test-Retest Reliability

Suppose researchers give a test to 20 individuals and then give the same type of test one month later to the

same 20 individuals.

Their scores are shown below:

	Test #1	Test #2
Individual #1	65	66
Individual #2	69	78
Individual #3	71	70
Individual #4	72	74
Individual #5	72	79
Individual #6	74	81
Individual #7	75	88
Individual #8	78	91
Individual #9	81	84
Individual #10	83	84
Individual #11	83	84
Individual #12	84	88
Individual #13	84	90
Individual #14	87	81
Individual #15	88	85
Individual #16	88	92
Individual #17	89	90
Individual #18	93	93
Individual #19	94	96
Individual #20	99	95

We can use a to find that the Pearson Correlation Coefficient between the two sets of scores is 0.836.

Since this correlation is greater than 0.80, researchers could conclude that the test has good test-retest reliability.

In other words, the test produces reliable results that can be replicated at different points in time.

Potential Bias in Test-Retest Reliability

1. Practice Effect

A practice effect occurs when participants simply gets better at some test due to practice. This means they're likely to show better results during later tests because they've had time to practice and improve.

The way to prevent this type of bias is to give individuals tests that are of equal difficulty but have a different variety of questions so that they can't memorize the answers to the types of questions asked on the first test.

2. Fatigue Effect

A fatigue effect occurs when participants gets worse at some test because they get mentally drained or fatigued from taking previous tests.

The way to prevent this type of bias is to provide plenty of time in between tests (ideally weeks or even months)

so that participants are fresh when taking both tests.

3. Differences in Conditions

When participants take the two tests under different conditions (i.e. different lighting, different time of day, different time allowed to complete the test, etc.) it's possible that they score differently on the tests simply due to differences in the testing environment.

The way to prevent this type of bias is to ensure that participants take both tests under identical conditions, i.e. during the same time of day, with the same general lighting and environment, and given the same amount of time to complete the test.