Reliability and validity

In the social sciences and psychology, testing internal reliability is essentially a matter of how consistent the results are. If the results for two sets of questions are similar, then the internal consistency is high. If they are different, then the internal consistency is low. This is because internal consistency is a measure of how well the different items of the test contribute to the overall score. If the results are consistently high, then the test is said to be reliable. If the results are consistently low, then the test is said to be unreliable.

When we talk about instruments, it does not necessarily mean a physical instrument, such as a scale or a computer. An instrument can also be a test or a survey. For example, a personality test or a psychological assessment tool. The principle is simple: if several expert raters all agree on a performance rating, that rating is reliable. If they disagree, then the rating is unreliable.

For most experiments of natural phenomena, results follow a normal distribution and there is a high degree of correlation between repeat scores. If the scores are 100, 111, 132 and 150, then the reliability and validity are also low. If the scores are 100, 102, 105 and 108, then the reliability and validity are high.

An example is clinical psychology role play examinations, where students are rated on their performance. The principle is simple: if several expert raters all agree on a performance rating, that rating is reliable. If they disagree, then the rating is unreliable. This is why, in order to ensure that there is a high degree of correlation between repeat scores, there will always be some disparities in these disciplines. For example, a test that routinely underestimates IQ by two points can be as useful as a more valid test since the difference is only two points. The key is to ensure that your results are as reliable as possible.

Another way of looking at reliability is by considering it as a way to maximize the inherent variability. If you use three or four experts and each gives a different score, then the overall reliability is low. If you use one expert and get exactly the same score, then the overall reliability is high.

Reliability and validity are two important concepts that are often confused. Reliability is concerned with the consistency of a test, while validity is concerned with the accuracy of a test. A test can be reliable but not valid, or valid but not reliable. The difference is best described with an example:

If the researcher's test delivers a consistent score of 118, then that's pretty close, and the test is reliable. However, if the scores are 100, 111, 132 and 150, then the reliability is low. The test is not measuring what it's supposed to, and it's doing it inconsistently. The test is not valid.

If the test consistently delivers scores of 135, and the candidate's true IQ is 120, the test is valid but not reliable. The test is measuring what it's supposed to, but it's doing it inconsistently. The test is not reliable.

If there is a high internal consistency, i.e. the results for the two sets of questions are similar, then the reliability is high. If you use three or four experts and each gives a different score, then the overall reliability is low. If you use one expert and get exactly the same score, then the overall reliability is high.

Reliability and validity are two important concepts that are often confused. Reliability is concerned with the consistency of a test, while validity is concerned with the accuracy of a test. A test can be reliable but not valid, or valid but not reliable. The difference is best described with an example:

If the researcher's test delivers a consistent score of 118, then that's pretty close, and the test is reliable. However, if the scores are 100, 111, 132 and 150, then the reliability is low. The test is not measuring what it's supposed to, and it's doing it inconsistently. The test is not valid.

If the test consistently delivers scores of 135, and the candidate's true IQ is 120, the test is valid but not reliable. The test is measuring what it's supposed to, but it's doing it inconsistently. The test is not reliable.

If there is a high internal consistency, i.e. the results for the two sets of questions are similar, then the reliability is high. If you use three or four experts and each gives a different score, then the overall reliability is low. If you use one expert and get exactly the same score, then the overall reliability is high.

Reliability and validity are two important concepts that are often confused. Reliability is concerned with the consistency of a test, while validity is concerned with the accuracy of a test. A test can be reliable but not valid, or valid but not reliable. The difference is best described with an example:

If the researcher's test delivers a consistent score of 118, then that's pretty close, and the test is reliable. However, if the scores are 100, 111, 132 and 150, then the reliability is low. The test is not measuring what it's supposed to, and it's doing it inconsistently. The test is not valid.

If the test consistently delivers scores of 135, and the candidate's true IQ is 120, the test is valid but not reliable. The test is measuring what it's supposed to, but it's doing it inconsistently. The test is not reliable.

If there is a high internal consistency, i.e. the results for the two sets of questions are similar, then the reliability is high. If you use three or four experts and each gives a different score, then the overall reliability is low. If you use one expert and get exactly the same score, then the overall reliability is high.