

A Quick and Dirty Guide to Validity & Reliability

Choosing the right assessment for selecting or developing employees can make or break the success of a talent initiative. Why bother using assessments that don't predict performance, or that fail to resonate with your business leaders?

When deciding on the right assessment for your valuable talent, pay attention to the scientific rigor with which the instruments have been tested. Any good tool should have concrete data demonstrating its validity and reliability. Validity and reliability can tell you two general things: 1) that the assessment is measuring what you want it to, and 2) that it will reliably assess the same thing each time - ensuring that the results you get aren't a one-off.

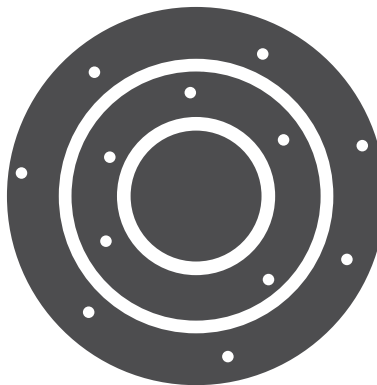
An easy way to think about this concept is with a bullseye metaphor:
The very center of the bullseye is exactly what you want to assess.

RELIABLE BUT NOT VALID



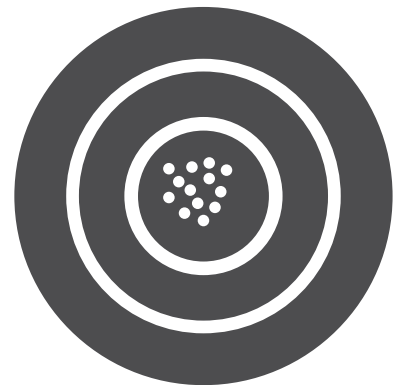
Reliable but not valid means that you are consistently testing the same thing over and over again, but it's not testing what you want to test.

VALID BUT NOT RELIABLE



Valid but not reliable means that the average scores align with the goals of the test, but individual scores are inconsistent.

RELIABLE *AND* VALID



Both reliable and valid means that the test will consistently measure what it is supposed to over a period of time - it's consistently hitting the bullseye.

What is Validity?

Validity refers to the accuracy of the assessment. In essence, does it measure what it is supposed to measure? While there are several types of validity to pay attention to, the most important for our purposes is predictive validity.

Predictive validity tells us how accurate a tool is at predicting a certain outcome. In the case of personality assessments, a good tool will be able to predict how well someone will perform their job. Validity is typically measured with a coefficient between 0 and 1¹ (called the Pearson correlation coefficient). The closer to one, the more accurate the predictive power of the test. The predictive validity of the Hogan Personality Inventory (HPI) is .29 for predicting performance across job families. However, when the HPI is combined with the Hogan Development Survey (HDS) and Motives, Values, and Preferences Inventory (MVPI), that number jumps to .54. While this may not seem very high, a good comparison is to look at the validity for something completely unrelated. For example, the predictive validity of ibuprofen for pain reduction is only .14. For another, more closely-related example, the correlation between structured job interviews and job performance is .18. There are many ways of measuring validity, some more useful than others. Any assessment provider worth their salt should be able to provide you with evidence of validity. If they don't, it's worth considering why not.

What is Reliability?

Reliability, on the other hand, refers to the consistency of the test. The reliability of an assessment can be evaluated in two broad ways: 1) internal consistency, and 2) test-retest reliability.

Test-retest reliability is a measure of consistency of responses over time. In other words, are people responding to questions the same way each time they take the test? Inconsistent responses can indicate that assessments results are not actually measuring personality, which should be relatively stable over time. Test-retest reliability uses a correlation of scores (again, using the Pearson coefficient) from a first assessment and a second assessment sometime later. For Hogan, the short-term test-retest reliability is .81 for the HPI, .70 for the HDS, and .79 for the MVPI.

Internal consistency relates to the questions that are used in each assessment. Test takers will notice that many questions appear to measure the same thing. This is on purpose. Asking a question in a few different ways helps us to ensure that we are getting an accurate measurement of the concept.

Like validity, reliability scores are also measured between 0 and 1 (this time with a coefficient called Cronbach's alpha). The closer to 1, the higher the reliability. The average internal consistency for the HPI scales is .76, .71 for the HDS, and .76 for the MVPI.

¹Absolute value. Scores between -1 and 0 indicate a negative correlation.

The important thing to note is that there is no *one* right way to measure reliability or validity. In fact, assessment publishers should constantly be monitoring their products to ensure they maintain the accuracy that they claim. Hogan Assessments are far above industry standards with continual evaluation of our own assessments. We are partial though, and we encourage you to seek out this information with any assessment system you choose.

Hogan Assessments have appeared in over 400 peer-reviewed publications to ensure that our tests are hitting the bullseye. We invite you to contact us for more information on the validity and reliability of Hogan Assessments at info@peterberry.com.au or +61 2 8918 0888.