

**A Quick Guide to**

# Validity and 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?**

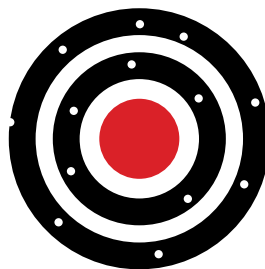
When deciding on the right assessment for your valuable talent, pay attention to the scientific rigor with which the instruments have been tested. Any reputable tool should have concrete data demonstrating its validity and reliability. Validity and reliability can tell you two general things: (1) that the assessment is measuring what you want it to measure, and (2) that the assessment 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.**



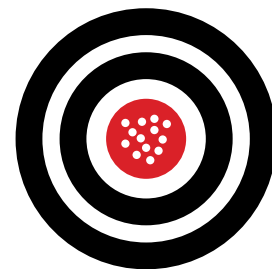
**▶ Reliable but not valid**

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



**▶ Valid but not reliable**

means that the average scores align with the goals of the test, but individual scores are inconsistent.



**▶ Reliable and valid**

means that the test will measure what it is supposed to measure over a period of time, consistently hitting the bull's-eye.



## ► 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, an effective tool will be able to predict how well someone will perform their job. Validity is typically measured with a coefficient between 0 and 1<sup>1</sup>. This is called the Pearson correlation coefficient. The closer the coefficient is to 1, 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 the 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 a more closely related example, the correlation between structured job interviews and job performance is .18. There are many ways of measuring validity, some of which are 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.

**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. At Hogan Assessments, we go far above industry standards with continual evaluation of our own assessments. But we are partial, of course, so we encourage you to seek out this information with any assessment system you choose.**

**Hogan's core assessments have appeared in more than 400 peer-reviewed publications to ensure that our tests are hitting the bull's-eye. We invite you to contact us for more information on the validity and reliability of the Hogan assessments at [info@hoganassessments.com](mailto:info@hoganassessments.com) or +1.918.749.0632.**

## ► 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 the 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 then 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 regards 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 is .76 for the HPI scales, .71 for the HDS, and .76 for the MVPI.

<sup>1</sup> Absolute value. Scores between -1 and 0 indicate a negative correlation.