With so many test publishers, how do I choose the right psychometric test?

Psychometrics Knowledge · https://psyasia.com/kb/with-so-many-test-publishers-how-do-i-choose-the-right-psychometric-test/

Human resource professionals are unlikely to need any convincing that the use of psychometric tests as an aid to employee selection and development is probably at an all time high. The increase in the use of aptitude and personality tests in the workplace is a positive thing provided the tests are chosen and used properly. This article discusses what decision-makers should look for in order to be confident they are making the right test choice.

The Hong Kong website of an employee testing system that is marketed worldwide claims:

"Really, what is the most effective way to evaluate the reliability and validity of any assessment tests so to help us to know exactly how to find the right productive people with certainty and predictability without any catastrophe in hiring any wrong people who simply look good?"

"The most workable and effective answer of the above questions is simply to TEST THE PEOPLE YOU KNOW VERY WELL; then you know which assessment test can be valid and reliable to use!"

This perspective is fundamentally flawed. It appears that the person who wrote it has no more understanding of tests than the average HR executive who wants to understand more and is looking to this article to assist them!

Many laypersons might assume that you can assess the validity of a test by self-completing it and/or asking somebody they know well to do likewise. Obviously, the assumption is that we know ourselves well and so if the test report provides an accurate reflection of the self that we know, it "must" be valid.

However, research shows us that individuals make flawed assessments of test reports.

In one research study, human resource professionals attending a conference were asked to complete a personality test. Following this, they were given a randomly generated narrative report. They were

NOT told that it had been randomly generated and were asked to evaluate its accuracy. 90% of the respondents agreed that the report was either amazingly accurate or very accurate (remember, the report was randomly generated).

It's partly for reasons such as the above that various worldwide psychological societies and academics have suggested that we assess at least 4 types of validity when evaluating tests. Validity refers to whether or not the test is fit for purpose (i.e., does it measure what it is supposed to measure or can it predict something that is meaningful, such as performance?).

Let's have a look at these important aspects of test validity:

Face Validity: Here, we simply ask if the questions in the test look like they are measuring what the test purports to measure. If I claim my test assesses numerical reasoning and you don't see any numerical data in the questions, you would doubt it has face validity. Assessing this type of validity is somewhat subjective and so it is considered to be the lowest level of validity.

Content Validity: We need to know whether the test questions are sufficiently representative of all of the possible questions that could assess the construct we are interested in. For example, if we want to assess conscientiousness, but my test simply asks questions that relate to your preference for following rules (only one aspect of conscientiousness), my test is unlikely to have content validity.

Construct Validity: You may have decided that your face-to-face salesperson must have a high level of self-confidence. If you are considering using an assessment of self-confidence to assist in your hiring decision, you'll need to evaluate whether the test really does assess the construct of self-confidence that it claims to measure. The best way to do this is to look in the publisher's manual for the test and find evidence that the publisher has correlated scores on this test with scores on established tests of the same construct. This aspect of validity is cited as one of the two most important. It is however somewhat technical as numbers are involved. It is better understood following training in the test or psychometric assessment generally.

Criterion Validity: This evidence is less easy to obtain than construct validity evidence, however it is also cited as one of the two most important areas of validity. Here we need to link scores on our test with performance. So, to take the above example again, one would expect scores on self-confidence to predict face-to-face sales performance. If they do, our test has criterion-validity. Again, the HR professional would look to the publisher's manual for evidence rather than carrying out the study themselves.

So, in terms of validity at least, evaluating and choosing the right test is a lot more complex than simply completing the test yourself or handing it to your colleague! Now we turn to reliability.

Reliability refers to the consistency with which a test assesses the construct of interest. Simply put, if I were to test you today and you scored 6 and tomorrow you scored 12, ignoring practice effects, we might suggest there is something wrong with the test! A more practical example would be the faith that you might place in a tape measure...if you measure the length of a table today and tomorrow and get different results, you know something is wrong. The measurement is inconsistent and so it is not reliable.

Reliability is vital for a test because if a test lacks consistency of measurement it can never be valid! No test is 100% reliable, just like no method of assessment is 100% reliable. Factors related to the test itself (such as ambiguous questions), the respondent (such as mood or exposure to tests) and the testing environment (such as noise and heat) can all impact upon the reliability of a test.

The website cited above stated that, as with validity, reliability is best assessed by having somebody you know complete the test! In fact, reliability is typically assessed by using the results of a sizable group of people, not just one or two people.

As with validity, there are a number of forms of reliability. Among them, internal consistency assesses the extent to which each question in the test is related to the overall scale score, whilst test-retest assesses the consistency of test scores over time. Reliability information should also be found in the publisher's manual. If it is missing or inadequate, it raises serious doubts about the integrity of the test.

Most psychometric tests that are used in selection require the comparison of the candidate's results to a group of similar others. This is how the score is made meaningful. If I told you I scored 7/20 on extraversion, this would mean very little to you. You might ask me how other people who took the test scored. It is therefore important that a test has been standardised on suitable groups of people - often referred to as a norm group.

One of our clients reported to us that she contacted the publisher making the claims above by email. She asked about reliability and validity of the test as well as whether or not local norms were available. She never received a reply, despite 3 reminder emails!

Additionally, if a test developed in one country is taken to another, it must go through a lengthy process of translation, validation and reliability checks. Many people do not realise this and assume that a test can easily be transported from one country to another just by taking it to a professional translator. This is not true!

As the use of psychometric tests in selection and development continues to soar, the human resource professional will need an understanding of how to evaluate tests.

In Asia in particular, we are noticing an influx of test publishers and distributors. However they are not always reputable and many do not have psychologists in the business at all! We have even heard of interested parties being told by publishers that validity information is protected and not available.

Worse still, one provider who has fairly recently headquartered in Hong Kong and is expanding throughout Asia claims that the founder has a PhD from a US university, yet when a client of ours contacted the university, he was informed that they have no record of the founder's PhD!

The publisher referred to at the top of this article, states:

"Instead of using "years" to really know the person followed with all kinds of risk, you can depend on our test to instantly know the person."

Whilst well designed, tested and validated assessments do provide extensive information on respondents that cannot be reliably and validly obtained using other less scientific methods, no reputable test publisher or distributor will claim their test can assist you to "instantly know the person".

In fact, it is imperative to schedule a feedback session with your respondent following personality testing to ensure that the profile you have for your respondent is valid. These sessions aim to elicit behavioural evidence from the respondent. Even after this exercise, the test user does not "know the person". They will rather have a good understanding of the individual which will assist them in their selection and placement decision.

Psychometric tests are thus useful tools in selection and development. They have their limitations however. With the growth of the industry and the adoption of test businesses by non-psychologists, it is in the interest of the test purchaser/HR professional to ensure they are suitably keyed up on how to evaluate the tests being marketed to them rather than blindly accepting strongly marketed but non-supported claims about tests.

Note on author's authority in this area:

The author is a doctoral-level registered organisational psychologist (Australia and Hong Kong). He has been delivering training in psychometric assessment that leads to the British Psychological Society's Certificates of Competence in Occupational Testing for over 20 years. His research in psychometric test validity has been recognised by the British Psychological Society with an award for Scientific Contribution to Occupational Psychology.

He has published his psychometric validation research in peer-reviewed international journals and he has reviewed related papers for the Society for Industrial & Organizational Psychology (USA).

His MSc research studied the link between personality and well-being/stress disorders in ambulance crews, whilst his PhD research investigated the validity of personality theory and questionnaires for the prediction of work performance in the People's Republic of China, Hong Kong SAR, Singapore, Australia, New Zealand and the United Kingdom.

He has significant experience in testing, working with multi-national companies worldwide as well as working with the Governments of Hong Kong SAR, Macau SAR, Malaysia, Singapore and the UAE.