

What sort of test should I use?

Nick Smith from EAM Consulting looks at the key differences between criterion and norm-referenced tests and when to use which type.

It is important to select the appropriate type of test for the knowledge or skill to be tested and to ensure that all of the stakeholders (candidates, trainers, line managers, senior managers and so on) know which type of test is being used and why. Only then can testing be fully effective.

So which sort of test should be used? The glib answer is that this depends on what you want to achieve or measure. If the answer to that question is “the competence of the individual” then we need to be clear which of the two main models of competence we are using – is it the Developmental Model or the Assessment Model?

“There is no reason why a particular test might not be used either for NRT or CRT

The Developmental Model looks at the abilities of high performers to identify the competencies (attributes) that make them successful. This approach is often found in appraisal systems using competence matrices where higher levels of competence are required for higher grade roles. These competencies are norm referenced.

By contrast the Assessment Model looks at the expectations of workplace performance and Competence is defined with regard of specific tasks in specific circumstances. There are no grades – candidates either can or cannot perform each task to the stated standard and so are either ‘Competent’ or ‘Not yet Competent’. These competencies are criterion referenced.

To illustrate the difference between the models, let’s look at an example. Mike and Andy both need to write to clients. The letters will need to set out clearly, in plain English, the matters at hand as well as meet certain compliance requirements.

Mike is a particularly skilled wordsmith and has attained level 4 of 5 on the firm’s competency matrix for communication skills. His letter sets out the matters at hand in clear and simple terms that are easy to read and understand. However, he omits one of the points required for compliance from his letter.

Andy is not such a skilled wordsmith and has only attained level 2 for communication skills. His letter does cover all of the compliance requirements but does not read as well as Mike’s, although it still sets out the matters at hand unambiguously.

I would suggest that whilst Mike has a high level

of competency in communication skills he is ‘Not yet Competent’ with regard to this specific letter writing task. Conversely Andy is ‘Competent’ at this task.

So which type of test should we use for this letter writing task?

If we decide that we need people to be ‘Competent’ at this task (they will meet the compliance requirements) then we need a test that will sort the candidates between those who can perform the task correctly and those who cannot. This is called criterion-referenced testing (CRT), because the candidates are tested against the criterion of correct performance. They could all pass or all fail.

If we decide that we need people with a high level of competency then we need a test that will sort out the more able candidates from the less able. The test should be able to discriminate between individual performances on the task being tested. This is called norm-referenced testing (NRT) because the candidates are tested against each other to determine who performs the task the best, and who performs the worst. A norm is established by analysing the test results; from this norm the tester can determine who is above or below the norm. These tests can also determine the strongest and weakest candidates.

There is no reason why a particular test might not be used either for NRT or CRT. In our letter writing example a case study requiring a letter to be drafted could be used as the test for either a NRT or CRT as it will be the marking scheme that will differ between the two.

Tests vary in their cost or resources. Generally speaking CRT is cheaper as the marking is typically a simple yes/no decision. NRT is more expensive as here judgement is required across the whole ability range. Finally the “both at once test” is very expensive because they are characterised by elaborate marking guides and final reviews to decide from mark sheets who should pass and who should fail *as well as* who is best and who is worst.

I would suggest that:

- ❑ CRT is used on “essential tasks” to decide whether a candidate passes or fails
- ❑ NRT is used on “desirable task features” to determine the order of merit

The big danger of the “both at once” test is that candidates may pass because good marks in one area mask poor marks in another. For example if our letter writing case study is marked for both use of English and the compliance requirements with a simple overall pass mark then Mike’s wordsmith skills may mean that he attains the pass mark even though his letter did not meet all of the compliance requirements.

CRT – NRT quick reference guide

| Type | Characteristics | Advantages | Disadvantages | Uses |
|------------|---|---|--|--|
| CRT | Performance oriented. Predetermined standard. Discriminates correct from incorrect performance. Pass – Fail decision. No attempt to discriminate between weak and strong candidates – only P / F matters. | Ensures correct performance by all who pass. Provides information on the effectiveness of the training. Directs candidates towards specific areas to be mastered. | Often resource intensive as all items may be tested. Does not identify strong from weak performers. | Evaluate ability to perform the task / job. Assess progress in training. Determine the need for training. Evaluate training “Cut off” for further training. |
| NRT | Usually represents a broad range of content. Grading is used to discriminate between candidates. Easy and difficult test items are used to help discriminate. Spread of scores is required. | Identifies strong from weak performers. Popular. Common experience of most people when at school. | Often does NOT indicate whether or not the candidate can perform the task. Often requires testing of non-essential or even trivial information to produce sufficient discrimination between candidates. | Select candidates for further training or promotion. Assess training strengths / weaknesses. Produce order of merit to reward outstanding performance. Enables attention to be focused on the best and the worst. |

A cautionary tale

Some years ago a RAF training centre invested in an expensive aircraft cockpit simulator for both the training and testing of aircraft technicians in the conduct of pre-flight checks. Prior to the use of the simulator the technicians had to visualise the cockpit and then write down the checks to be carried out in order. The first time pass rate on the test was around 70% and the RAF units complained that even though the technicians had passed the test before arriving at their units they could not carry out a pre-flight check correctly. The use of the simulator meant that the technicians could practise carrying out the checks in a cockpit rather than having to imagine them. Testing was carried out by completing the checks in the simulator rather than having to write about them. As a consequence the first time pass rate rose to over 90% and the feedback from the RAF units was that the new technicians were arriving able to carry out the checks correctly. The trainers were rightly pleased with these results and so were also rather shocked when their Station Commander expressed his concern that something was going wrong – the increased pass rate must mean that the test had become too easy and the trainers were to make it more difficult so that the pass rate came down again!

The problem was that whilst the test was, quite rightly, a CRT the Station Commander was interpreting the results on the assumption that the test was a NRT. Given that most of the examinations we have undertaken for qualifications (starting with those we took at school) are NRT then it should not be a surprise that many people assume that tests are NRT.



Hence, not only is it important to select the appropriate type of test but it is also vital to ensure that all of the stakeholders (candidates, trainers, line managers, senior managers and so on) know which type of test is being used and why if testing is to be effective.

Nick Smith is a specialist in the evaluation, assessment and measurement of training
 nick@eam-consulting.co.uk
 07775 586235
 www.eam-consulting.co.uk