



Brushing up on assessment

# Understanding tests

3

One of ten guides for school practitioners and those looking to brush up on their assessment knowledge

In education, the term assessment refers to the wide variety of methods that educators use to evaluate and measure pupil knowledge, understanding and learning. These methods can range from informal teacher observations to standardised tests; good assessment practice recognises and values many types of evidence, both formal and informal.

While informal assessment (such as strategic questioning, classroom quizzes and observations) is useful in evaluating ongoing learning on a daily basis, formal assessments (such as tests) can provide useful evidence about what pupils know and can do at a particular point in time (e.g. before or after a period of learning).

This guide offers an overview of why schools use formal and often standardised tests, how they are developed and what to bear in mind when selecting tests for your school.

## REMEMBER:

All tests are assessments, but not all assessments are tests.

For more on the different types of assessment, see '*Starting out in assessment*', the first guide in this series, available at ([www.nfer.ac.uk/brush-up-assessment](http://www.nfer.ac.uk/brush-up-assessment))

## Why schools use tests?



Used alongside outcomes from other assessments, tests help give a more complete picture of pupils' attainment. They can be particularly useful for highlighting strengths and weaknesses of individual pupils, or across classes and year groups, providing diagnostic information to inform ongoing teaching. At the beginning of term, or a period of new learning, tests can be useful in providing a baseline measure of knowledge and ability to compare against. Tests can also provide valuable summative data at the end of a period of learning, enabling schools to monitor attainment and progress at individual, class, year group or school level.

If tests are taken under controlled, standardised conditions, the results of pupils in different circumstances can be directly compared. Some tests provide age-standardised scores, meaning that age is taken into account when comparing the test performance of different pupils.

In the administration guidance of good tests, the purpose of the assessment is made clear. This includes who the test is intended for, what is being assessed and how the results should be interpreted. Good tests are also designed to be accessible to those taking them, ensuring that the format, style and content are age-appropriate and fair for the intended test takers. These features contribute to the test being a valid assessment.

If tests are taken under controlled, standardised conditions, the results of pupils in different circumstances can be directly compared.

## How are tests developed?



Robust tests are rigorously developed, following a sequence of stages. There are five main elements in most test development cycles:

1

### TEST SPECIFICATION

During this stage consideration is given to the purpose of the test and the target test takers. This informs a specification of the required test characteristics, outputs and appropriate methodology for the development of the assessment.

2

### ITEM (QUESTION) WRITING

Test questions are then written and reviewed in order to map to the relevant curriculum or to adequately represent the particular skills, knowledge or understanding the test is intended to assess.

3

### TRIALLING

Questions are trialled in schools to establish how pupils interact with them, to inform the development of the mark schemes and to identify any aspects that could be improved or amended. Feedback from the informal trial is used to refine test questions. The scale of this element can vary but is an important component in rigorous test development.

4

## STANDARDISATION

Once questions have been finalised, standardised tests are trialled on a large statistically representative sample of pupils prior to the test becoming live, in order to provide national benchmarks. Rigorous analysis of the final trial ensures that the test is technically sound, enabling conclusions to be drawn with confidence.

5

## STANDARD SETTING

This is a process involving both qualitative and quantitative data that is used to define the threshold levels of achievement (cut scores or grade boundaries) or the required proficiency on an assessment. For example, to establish what raw score on the test corresponds to the 'expected standard' for a pupil.

## Selecting a test

Some tests are statutory (e.g. national curriculum tests in England and national reading and numeracy tests in Wales) while others are optional (published tests). Statutory national tests are developed rigorously as outlined above, as are some published tests, including NFER's KS1 and KS2 assessment range. However not all providers follow this rigorous process of review, trial and refinement.

When selecting a test for your school, take account of information provided by the developer about the development process, as well as the content, to help you evaluate the **validity** of the test including whether it provides appropriate coverage of the specified domain e.g. the curriculum. Take account also of information about administering the test and interpreting the results. Most technically sound tests will offer some advice on administering the test in a specific way to ensure rigour is maintained, and also to give comparability with other schools using it. This is particularly true for **standardised** tests.

### Key questions to ask when choosing tests:

- Do the tests have full and accurate coverage of the curriculum?
- Have the tests been developed by assessment experts with input from teachers and other relevant third parties?
- Are the tests valid i.e. are they appropriate for the purposes for which you require them?
- Have the tests been standardised with a large number of pupils from a nationally representative sample of schools?
- Do the tests come with teacher guidance and mark schemes?
- What results will the tests provide?



Here are some key questions to ask before you purchase tests and crucially, why you should ask them:

### HOW WELL DO THE TESTS COVER THE SPECIFIED ASSESSMENT DOMAIN?

If you want to be able to say how well pupils are performing in reading and maths as defined in the national curriculum, for example, you need to ensure that the tests you use provide appropriate coverage of the current national curriculum. You will want to be sure that your chosen test is assessing what has been learnt from the curriculum, rather than just assessing a generic skill. In subjects where the programme of study covers several years, you also need to be confident that there is appropriate coverage in the test for each year.

### WHO HAS DEVELOPED THE TESTS?

A well-developed test will have been designed and reviewed at various stages by many interested parties. These include assessment experts with the technical knowledge required for robust test development, and teachers who have a solid understanding of the curriculum and indicators of pupil ability. This involvement is necessary to develop a test that delivers reliable outcomes.

### HAVE THE ASSESSMENTS BEEN STANDARDISED?

A crucial element of robust assessment is standardisation. A standardised test is one that has been trialled with a nationally representative sample of pupils. The outcomes from standardised tests enable you to benchmark your pupils against the national average and to compare performance on different tests. It might surprise you to learn that some published tests have never been trialled in schools.

### HOW LARGE WAS THE NATIONALLY REPRESENTATIVE SAMPLE?

The size of the pupil sample is important if you wish to benchmark your pupils against attainment nationally. Usually this requires a minimum of around 1,200 pupils per test across a range of schools representing different regions, types and performance bands. Larger samples will give more accurate standardised scores, which is why NFER Tests for Years 1-5 were standardised with over 60,000 pupils.

### ARE THE RESULTS RELIABLE?

Reliability has a very specific meaning in test development, but for classroom assessment purposes you need to be sure that the tests are administered and marked in a consistent way. This means that the administration guidelines must be clear. In addition, assessments which have been properly developed and trialled will have mark schemes which reflect what teachers actually see in children's responses. Without clear administration guidelines and properly developed mark schemes, you run the risk of obtaining unreliable results.



Reliability has a very specific meaning in test development, but for classroom assessment purposes you need to be sure that the tests are administered and marked in a consistent way.

## WHAT RESULTS DOES THE TEST PROVIDE?

The most straightforward outcome when a pupil takes a test is their raw score, but this doesn't enable you to compare results across tests or over time. Standardised scores enable you to do this. NFER Tests offer both standardised scores and age-standardised scores, which can be particularly useful for putting the performance of younger pupils into context. In addition, NFER's summer term tests identify how children are performing against curricular age-related expectations.

## IS IT VALID TO USE THE RESULTS IN THE WAY YOU INTEND TO?

What determines the validity of an assessment is how well it supports the uses to which the results are going to be put. If you want to use a test to identify the pupils who may benefit from a phonics intervention, for example, then the test specification (of the test content) needs to cover phonics rather than focusing on reading comprehension. A test may be a valid assessment for one purpose but not for another.

## Found this useful?

This guide is part of NFER Classroom's *'Brushing up on assessment'* series, a collection of free guides providing those looking to build their knowledge with an overview of key elements in assessment.

Sign up to our assessment newsletter ([www.nfer.ac.uk/assessment-newsletter](http://www.nfer.ac.uk/assessment-newsletter)) to be the first to receive each guide as soon as it is published, as well as other exclusive assessment-focused content direct to your inbox.

Visit [www.nfer.ac.uk/brush-up-assessment](http://www.nfer.ac.uk/brush-up-assessment) to view all the guides in this series.

## Looking for assessment materials?

# NFER Tests for years 1-5

Thousands of schools use NFER Tests to support attainment and progress monitoring. These termly assessments are:

Trialled with 60,000 children to ensure age-appropriate materials and meaningful data.

Written by a team with assessment expertise and teaching backgrounds.

Supported by an easy-to-use online analysis tool.

Visit [www.nfer.ac.uk/tests](http://www.nfer.ac.uk/tests) to find out more.