**Helping teachers make the most of classroom assessment**

**Bethan Burge**

Centre for International Comparisons, National Foundation for Educational Research

[b.burge@nfer.ac.uk](mailto:b.burge@nfer.ac.uk)

**Abstract**

Classroom assessment is a vital part of teaching and learning, helping teachers to track the progress of their students and informing their decisions about next steps. Many teachers do not have time, or lack the knowledge, skills or confidence required, to develop their own classroom assessments. As a result they use commercially available tests to assess their students. As test developers it is our responsibility to help to ensure that these assessments are well developed to provide reliable evidence that can be used to support teacher judgements. It is also important that there is guidance for teachers on how to interpret the test data in a meaningful and robust way. It may also be valuable to provide information and guidance to teachers on how the data from assessments can be used to identify the next steps that students need to take in their learning.

This paper will explore:

* the role of the test developer in making test results meaningful to teachers
* the importance of ensuring the test items reflect what is being taught in classrooms
* the categorisation of items to aid the analysis of student response patterns.

**Background**

The role of the test developer in *designing the assessment of learning outcomes to make a positive impact on individuals and institutions* will inevitably be influenced by the current system of assessment and accountability. This paper focuses on what test developers can do to help support teachers in England in their assessment of students. Namely how to ensure that the tests we develop can provide teachers with information that they can use in a formative way as well as providing evidence for summative teacher assessment judgements. Despite the increased autonomy offered to schools in England in the last few years, the system is still dominated by high-stakes testing and accountability. In 2010 the UK government passed the Academies Act this opened up the academies programme allowing all state-funded primary, secondary and special schools to apply to convert to academy status[[1]](#footnote-1). Since the legislation was passed a large number of schools have converted to academy status, as of May 2012, 1807 schools in England have opened as academies (this is a big increase from the 203 in May 2010).The Academies Act also allowed for Free Schools, these are new schools proposed by independent groups, which if approved by the Department for Education receive funding from central government. All state funded schools, even Academies and Free Schools which have moved away from local authority control are required to deliver a number of statutory assessments. In addition, teachers must use their knowledge of a student’s work over time to make a teacher assessment judgement on a student's progress and performance. Teachers are advised that these assessment judgements should not be based solely on test scores but should include evidence of performance in a variety of contexts collected over a period of time. These judgements need to be submitted to the government at the following stages:

* at the end of lower primary education (Key Stage 1) (for English, maths and science)
* at the end of upper primary education (Key Stage 2) (for English, maths and science) (along with test results in English and maths)
* at the end of lower secondary education (Key Stage 3) (for all core subjects).

This information needs to be provided for all students including those who are working at a low level.

Whilst these statutory assessments and teacher assessment judgements may not be considered high stakes for the students themselves, the stakes are high for teachers and schools as this data is used as a key accountability measure. Since 1997 the government in England has published performance tables which include test and examination results for each school. The performance tables for Key Stage 2 (the end of primary education) continue to report results for the statutory assessments in English and maths taken at the end of primary education. In recent years there has been a move away from reporting test results at the end of lower primary (Key stage 1) and lower secondary (Key Stage 3) education in favour of reporting teacher assessment judgements. However, as this information is still publicly available it can be used to judge schools and local authorities.

**Teachers’ use of assessments**

Assessment is an important feature of most education systems. Newfields (2006) reported research which showed teachers spend an estimated 10 to 50 per cent of their time on assessment related activities. International comparison surveys such as PISA collect detailed information about assessment. The PISA study gives an insight into the way in which assessments are used in schools across the world as well as the use that is made of the results. PISA 2009[[2]](#footnote-2) asked headteachers how often students at the end of secondary education were assessed using different forms of assessment. Headteachers in the UK reported that students are most frequently assessed using teacher developed tests, 57 per cent reported that students in their schools are assessed using these types of test between three and five times a year. Students are also assessed with commercially available standardised tests, although this happens less frequently with 57 per cent reporting that students are tested one or two times a year using this method. According to the PISA 2009 findings schools in the UK use assessment data for a variety of purposes. Namely, to inform parents about their child’s progress (99%); to monitor the school’s performance from year to year (96%); to group students for instructional purposes (94%); to identify aspects of instruction or the curriculum that could be improved (92%); and to compare the school to local or national performance (91%).

If schools are to inform parents of progress and group students for instruction, as is indicated by the PISA results, they are likely to require more detailed information than the raw scores from teacher developed tests alone will produce. Teachers may benefit from using other types of assessment that provide more robust data and are more appropriate for collecting evidence that can feed into summative teacher assessment judgements. Whilst test developers may understand the difference in the quality of the data obtained from one assessment compared to another, for example raw scores versus age standardised scores, this is not necessarily the case for teachers. Some research has shown that teachers may be lacking in what has been termed *assessment literacy*. Whilst there is not necessarily an agreed definition of the term assessment literacy a key feature is likely to be ‘understanding which assessment methods to use to gather dependable information about student achievement’ (Wang, *et al*., 2008).

**Developing teachers’ assessment literacy**

The ability to select an assessment that will give both robust and useful information is particularly important for teachers in England, where increasingly, teacher assessment judgements are reported publicly and used for accountability purposes. Although test developers may not be directly involved in training and developing teachers, there are ways in which they can help to inform teachers’ assessment literacy and provide teachers with richer assessment evidence through the design of assessments and the accompanying teacher guidance materials.

**Making test results meaningful to teachers**

Choosing the right assessment is vital if the results are to be meaningful and inform teacher assessment judgements. When a teacher is choosing an assessment their decision is likely to be informed by a number of considerations: which skill do they want to assess, who do they want to assess and what data will they get from the assessment.

It is important that test developers describe clearly what the test will actually assess. For example, if a teacher wants to assess their students’ reading ability they need to be aware that not all tests assess the same aspects of reading. In the case of a reading test, it is the responsibility of the test developer to make it clear exactly what aspects of reading will be assessed using that particular test, i.e. is it primarily a test of decoding or comprehension.

The teacher guidance for a test should also make clear who the test is intended for. It should describe the age group for whom the test has been developed and whether group or individual administration is more appropriate. Importantly there should be guidance as to whether the test can be adapted to make it more accessible without impacting on the reliability of the results and if so what aspects should be adapted to improve accessibility. If adaptations will lead to changes to the way that the test functions there should be guidance on what the likely impact will be on a student’s results. It is equally important that a teacher is aware of the limitations of a particular test. For example, if a test is designed as a group monitoring test it would not provide the level of detail needed for an individual diagnostic test.

It is important that a teacher understands the different scores that they will get from a test and how to interpret these in order to make informed judgements about a student’s ability. It is good practice to explain the benefits of reporting one type of test score over another in any teacher guidance material that accompanies an assessment. For example, acknowledging that age standardised scores are more accurate than raw scores because they take into account the age of the student taking the test and are more useful as they allow scores from two different tests to be compared meaningfully and allow for a teacher to compare their students with a large nationally representative sample of students. Including this information in the teacher guidance can help to inform teachers’ assessment literacy. Another important aspect of assessment literacy is an understanding of the accuracy of the results of an assessment as an indication of student ability. Teachers should be made aware of errors of measurement and confidence bands of the assessment they are administering so that they can make informed judgements about the accuracy of a student’s test results.

**Ensuring the test items reflect what is being taught in classrooms**

If results from a test are going to usefully contribute to teacher assessment judgements the test needs to assess what students are being taught in the classroom. This should be a primary consideration throughout the test development process. At the start of the process it is vital to establish what students are being taught; this may involve reviewing curriculum and policy documents, talking to teachers about their practice and what they think should be included to create an authentic assessment and also looking at the content of statutory assessments. The information gathered at this stage should feed into the development of the test and item specifications which will set out the aspects of the curriculum that are going to be tested, the technical aspects of the test (such as the length of test and time allowed) and the format of the test (for example, whether the test will include multiple-choice items). It is important teachers and students are involved early in the development process to check that the decisions that have been made about the coverage and content of the test and individual items are appropriate. At this stage the items should be piloted with small groups of students and the feedback from the students and their teachers used to refine the items and tests.

**Categorisation of items to aid the analysis of student response patterns**

The value of a total raw score as a piece of assessment evidence is fairly limited, as it only really enables a teacher to make a judgement about the relative performance of their students on a particular test. A raw score for a test does not provide a teacher with information about a student’s strengths or the skills they may need to develop further and therefore does not do much to inform the next steps for teaching and learning. In principle, it would be possible for a teacher to obtain this more detailed assessment information by examining completed test scripts, establishing which questions have been answered correctly and which were more difficult. Judgements about a student’s performance based on the analysis of response patterns, as described above, would also need to be informed by an understanding of which curriculum areas, concepts or topics were being assessed in each question.

It is the role of test developers to make assessments useful for teachers and categorising test items is one way to ensure that teachers, and indeed students, get more out of the assessments they use. During the test development process it is good practice to devise a test specification which outlines the proposed content of the test and in particular the balance of items needed to cover a curriculum area, concept or topic being tested. As questions are developed they should be categorised according to the requirements of the test specification so that during the test construction phase it is possible to establish whether the desired balance has been achieved across the test. Making this categorisation information available to teachers would enable them to more easily analyse student response patterns, providing useful assessment evidence that can be used to support teacher judgements. For example, when a student takes a reading comprehension test it is important for the teacher to know that a particular student is able to answer the items that require them to use retrieval skills but has more difficulty with the questions where they have to use inference and deduction to find the correct answer. This information can be used to inform decisions about appropriate learning activities that will help to develop these skills. However, it is important that teachers take into consideration the number of items that focus on the skill in question as this is likely to have an impact on the reliability of these judgements.

**Exploring student response patterns to** support **teacher assessment**

The following section gives examples from two NFER developed tests: New Group Reading Test (published by GL Assessment) and NFER Formative Assessment Service. The examples demonstrate how NFER test developers have used item categorisation and analysis of response patterns to provide teachers with assessment information that can feed into summative teacher assessment judgements and also helps teachers to plan next steps for their students’ learning.

**New Group Reading Test (NGRT)**

The New Group Reading Test (NGRT) is a short, easy-to-administer test of decoding and reading comprehension that can be administered to a whole class. It is intended to be used to assess the early reading skills and monitor students’ progress up to the age of 16. The NGRT has alternate forms to allow for re-testing at the end of educational phases to measure progress.

In the development of the third edition of the NGRT changes were made to both the content and the format of the test to reflect the growing importance of teacher assessment across the UK. Whilst the NGRT is still a multiple-choice test, the questions have been categorised to aid the analysis of student response patterns to provide useful assessment evidence that can support teacher judgements. In addition to providing evidence for summative teacher assessment judgements, this information can help teachers in their planning of targeted lessons or activities that will help develop a student’s reading ability further.

Below is an example of a categorisation table for one pair of the NGRT tests. Table 1 identifies how many of the questions in the passage comprehension section of the NGRT can be attributed to the different English National Curriculum reading Assessment Focuses (AFs). This tells teachers which aspects of the reading AFs will be assessed. The teacher is then well placed to decide whether this assessment will provide them with enough information or whether they will need to supplement this with additional activities or assessments.

**Table 1: Assessment Focus coverage for NGRT**

|  |  |  |  |
| --- | --- | --- | --- |
| **Test section** | **Item type** | **Number of items Test 3A** | **Number of items Test 3B** |
| Sentence completion | Sentence completion | 20 | 20 |
| Passage comprehension | Context comprehension | 12 | 13 |
| Retrieval | 4 | 6 |
| Inference and deduction | 8 | 8 |
| Organisation of texts | 2 | 2 |
| Writer’s use of language | 4 | 2 |
| Writer’s purpose and viewpoints | 2 | 1 |

The marking keys included in the teacher guidance give further information about the Assessment Focus categorisation for each question in the passage comprehension section. Table 2 shows the categorisation of the questions for one of the comprehension passages.

**Table 2: Item categorisation for comprehension passage**

|  |  |  |
| --- | --- | --- |
| **Item no.** | **Correct answer** | **Item type** |
| 1 | rewarding | Context |
| 2 | adventure | Context |
| 3 | desire | Context |
| 4 | to persuade you to join Youth Challenge | Writer’s purpose and viewpoints |
| 5 | to show you what each section is going to be about | Organisation of texts |
| 6 | You can customise your programme. | Retrieval |
| 7 | It will make you excited. | Writer’s use of language |
| 8 | sceptical | Inference and deduction |
| 9 | so you can get more information | Writer’s purpose and viewpoints |

Student case studies form part of the teacher guidance materials. These case studies exemplify the performance of students across the ability range. Researchers at NFER developed these case studies using test scripts from the standardisation process. Each case study includes:

* scores from the NGRT (raw score, standardised score and reading age)
* background information about the student
* analysis of response patterns for the two sections of the NGRT test
* next steps for teaching.

The case studies illustrate the way in which teachers can look in detail at an NGRT test script from an individual student in order to provide evidence for areas of strength or identify skills that may need to be developed through classroom activities or targeted support. The case studies also demonstrate how the NGRT can be used in conjunction with a more in-depth diagnostic test of reading to further explore a student’s performance. Whilst a teacher might not do this for every student the case study approach can be useful in interpreting unexpectedly low scores which cannot be easily explained without gathering more evidence of the nature of the reading difficulty.

Below are some extracts from one of the case studies included in the NGRT teacher manual.

**Figure 1: Extract from case study including scores on NGRT and background information**

**Dennis, 11 years and 9 months**

|  |  |  |
| --- | --- | --- |
| **NGRT Raw Score** | **NGRT Age Standardised Score** | **Reading Age** |
| 22 | 86 | 9:00 |

Dennis is in year 7 at a community college in the south west of England. His teacher estimates him to be working at level 3c. Dennis has a limited degree of Special Educational Needs support (School Action).

Overall, Dennis received 22 marks out of a total of 52 marks on the *NGRT* 3A, he has an age standardised score of 86 and is in the 18th percentile. This score indicates that his reading ability is lower than would be expected for his age compared to the national sample.

**Figure 2: Case study extract of the analysis of the response pattern for the passage comprehension section of the NGRT**

…However, whilst he was able to understand the overall purpose of the text, he was unable to assimilate the detail, and the majority of his mistakes were for questions that required him to make an inference from the text. For example, in ‘Walking the Line’, students have to infer the meaning of the ‘writhing mass’ in the sentence, ‘Looking down I couldn’t really see the crowd, just a swarm of black dots that blurred into a writhing mass’. Dennis thought that the writhing mass were birds. Birds are mentioned in the first paragraph, but their significance is made through the comparison between the author’s view (a tightrope walker), and the view held by birds in flight. A similar situation occurs in ‘The Eiffel Tower’, where students were asked what the Exposition Universelle celebrated, drawing upon the sentence, ‘…Exposition Universelle, a World’s Fair marking the centennial celebration of the French Revolution’. Dennis thought that the Exposition was a 100 year anniversary of the World’s Fair, instead of the French Revolution.

**Figure 3: Case study extract of next steps for teaching**

Dennis’ reading profile suggests that his reading skills are still heavily focused on decoding the text and as a result his comprehension skills are lagging behind. Dennis still needs regular practice to improve his reading speed and fluency. He has a fairly good understanding of grammar and word function, and is capable of recalling information from the text, but finds it harder to interpret meaning from longer text passages. It would benefit Dennis to spend further time discussing the text, with a partner or teaching assistant, explaining his answers and exploring alternative interpretations. Where possible, he should be encouraged to refer back to the text to provide evidence for these.

**NFER Formative Assessment Service**

The NFER Formative Assessment Service (NFER-FAS) is an online assessment for primary schools that can help teachers to assess the progress of individual students and provides teachers with information for a number of formative and summative assessment purposes. The type of assessment information ranges from basic scores on individual assessments to complex profiles of student performance in a particular curriculum area (there are NFER-FAS assessments in reading, mathematics and science). The NFER-FAS produces profiles for students based on their performance on one or more of the online assessments. These are not based on the total score a student achieves but take into account the pattern of responses across different kinds of questions, and how difficult the questions are. The profiles produced by NFER-FAS go beyond test scores to show patterns of understanding, partial understanding and misconception. The next section describes the main components of the profile reports and includes extracts from one of the science profile reports.

Each profile report includes several components, which are described below:

**Description**

This part of the report describes the pattern of performance that defines the profile. It complements the teacher’s own understanding of each student. It could be used to give focused feedback to students, as well as forming the basis for planning learning experiences. It includes the average scale score for students with that profile, which gives a broad indication of the level of attainment. However, profiles are not entirely dependent upon scores, as they describe what students can and cannot do, rather than just counting correct answers.



**Next steps for learning**

This part of the report provides teaching suggestions related to the strengths and weaknesses identified by the profile. The teaching suggestions include various classroom activities and differentiation, to support students of differing attainment and different strengths and weaknesses, as identified by the profiles.



**‘Child-speak’ version**

Here, the pattern of performance is described in terms that students should be able to understand (perhaps with help from the teacher). These are provided to help with self-assessment, to support assessment for learning. They may be used as a basis for discussing learning objectives, or for setting individual targets.

****

**Profiles and uncertainty**

An important aspect of assessment literacy is an understanding that however carefully an assessment is constructed an element of unreliability is likely to appear in the results they produce. It is important that any teacher guidance that accompanies an assessment explains this potential uncertainty to the teachers who are using the test so that they can take it into consideration when interpreting student performance. The teacher guidance that accompanies the NFER-FAS explains to teachers that the profile reported for each student is the ‘most probable’, calculated with an associated percentage probability. Teachers are able to check the percentage probability of a student’s assigned profile on the NFER-FAS website. If the assigned profile does not seem appropriate to the teacher based on their knowledge of the student, they are encouraged to check the probability for each of the other profiles.

**Conclusion**

Test developers are able, through the provision of detailed teacher guidance, to help teachers develop their assessment literacy. Explaining to teachers how to interpret the scores they obtain from an assessment as well as describing the limitations may help teachers to select the assessments that will give them results that are meaningful. It is also important that test developers consider how a test can be constructed to provide teachers with the most useful data. The examples above demonstrate that tests can be developed in a way that provides teachers with a wealth of information far beyond simple raw scores. This type of detailed assessment evidence is of much greater value as it can be used not only to strengthen summative teacher assessment judgements but also to inform decisions about targeted activities that will help students to develop.

**References**

National Foundation for Educational Research (2007). *Getting to Grips with Assessment: Primary*. Slough: NFER [online]. Available:

<http://www.nfer.ac.uk/publications/getting-to-grips-with-assessment-primary/>

National Foundation for Educational Research (2010). *New Group Reading Test:* *Teacher* *Guide*. Chiswick: GL Assessment.

National Foundation for Educational Research (2012). *NFER Formative Assessment Service: Teacher Guide.* Slough: NFER.

Newfields, T. (2006). ‘Teacher development and assessment literacy.’ Proceedings of the 5th Annual JATL Pan-SIG Conferenc*e* ‘Authentic Communication’, Tokai UniversityCollege of Marine Science, Shizuoka Japan, 13-14 May [online]. Available: <http://jalt.org/pansig/2006/PDF/Newfields.pdf> [3 September, 2012].

Sainsbury, M. (2009). ‘e-assessment in school: from innovation to integration.’ Presentation at the Tenth Annual AEA-Europe Conference ‘Innovation in Assessment to Meet Changing Needs’, Balzan, Malta, 5-7 November.

Wang, T. H., Wang, K. H. and Huang, S. C. (2008). ‘Designing a web-based assessment environment for improving pre-service teacher assessment literacy’, *Computers & Education*, **51**, 1, 448-462.

1. Academies are independent, state-funded schools, which receive their funding directly from central government, rather than through a local authority. They have more freedom than other state schools over their finances, curriculum, length of terms and school days and do not need to follow national pay and conditions for teachers. [↑](#footnote-ref-1)
2. The 2009 data was obtained using the *Interactive Data Selection* tool <http://pisa2009.acer.edu.au/interactive.php> [↑](#footnote-ref-2)