

**Strengthening teacher assessment in schools in England: the development and use of *Assessing Pupils' Progress***  
**Stephen Anwyll and Dr Sue Horner**  
**QCDA England**

## **Background**

England has a long and well-documented history of national assessment since the introduction of the first national curriculum in 1988 and the establishment of statutory tests and teacher assessment in the early 1990s. For more than a decade, all 7 year olds (end of key stage 1) were tested in reading, writing and mathematics and all 11 year olds (end of key stage 2) and 14 year olds (end of key stage 3) in reading writing, mathematics and science. At key stage 1, the tests were marked by each school and at key stages 2 and 3 they were marked externally. Teachers were also required to make teacher assessment judgements in speaking and listening, reading, writing, mathematics and science of the same pupils at the end of each of the three key stages. All judgements were related to national standards, set out in an eight-level scale for attainment targets within each subject.

Although test outcomes and teacher assessment judgements were both reported to parents and submitted nationally, it was the test result which came to have more status in the media, when school-level results were published for 11 and 14 years olds, and in the use of achievement data within the public inspection of schools by the Office for Standards in Education (Ofsted). At key stages 2 and 3, no significant national funding went into supporting the quality and accuracy of teacher assessment judgements and many teachers waited until test results were received before finalising their own assessment judgements. Regulatory activity has focused almost exclusively on the test arrangements and there has been no requirement for local authorities or schools to undertake any moderation activity to secure teacher assessment judgements at these two key stages.

## **Changes in national assessment policy**

Significant changes have been made to national policies over the last seven years. The most recent of these were announced by the government's *'Expert Group on Assessment'* which published its recommendations in May 2009. A brief summary of the main changes at each key stage is given below.

**Early Years Foundation Stage (0-5 years):** A national Foundation Stage Profile was developed to bring greater consistency to the assessment of children on entry to school around the age of five. This consisted of an observation schedule linked to scales relating to the Early Learning Goals of the six Areas of Learning in the Foundation Stage. The use of this Profile became statutory for all early years settings attended by children in the final year of the Foundation Stage (Reception year) in 20xx. It has subsequently become known as the Early Years Foundation Stage Profile and it is formally moderated by Local Authorities (there are 150 of these across England responsible for organising local educational provision). This

moderation takes the form of visits by experienced practitioners to a sample of local settings each year to look at assessment processes in use and to check a small sample of children. After concerns were expressed about the consistency with which scale descriptions were being interpreted, centrally organised and assessed moderation training is now being provided for representatives of all Local Authorities.

**Key Stage 1 (5-7 years):** After a pilot in 2003-04, statutory assessment at the end of key stage 1 was changed in 2004. Since then, only children's teacher assessment levels have been reported to parents and collected nationally. Schools are still obliged to make use of centrally-devised, national tests to inform their overall judgements but these can be used flexibly and schools have a choice of tests from which to select. Local Authorities retain the duty to moderate key stage 1 assessment but from 2004 the focus switched away from the correct use and marking of the tests towards scrutiny of the processes used for teacher assessment and the accuracy of teachers' judgements. As with the Early Years Foundation Stage Profile, moderation is undertaken with about 25% of schools each year and is usually carried out by a mixture of Local Authority advisory staff and experienced local teachers.

Some Local Authorities organise events where teachers, particularly those new to teaching at Year 2 (the final year of the key stage), can bring samples of work from their pupils in order to discuss and informally moderate judgements and there is a limited amount of cross-Authority moderation activity organised by members of the national assessment advisers' organisation (AAIA).

**Key Stage 2 (7-11 years):** national, confidential, externally-marked tests in reading, writing, mathematics and science have been in place for England's 11 year olds for almost twenty years. Annual publication of the national results attracts media attention each summer and individual school results, which are published by the press as 'league tables' in the autumn, get considerable attention from local estate agents who believe that parents will want to know where the 'best' schools can be found. These are high stakes tests and the results play a major part in the evidence used by inspection teams which visit schools at a few days' notice, generally once every four years, and publish their reports. Nationally, they have also been a key focus of educational investment with national targets set and major initiatives (such as the Primary National Strategy) set up to help achieve them.

Since the turn of the century, there has been an increasing shift of focus away from considering simply the percentage of pupils reaching a particular level by the end of the key stage to also looking at the progress made by children between the end of key stages 1 and 2. These so-called 'value-added' measures have become increasingly sophisticated, taking into account a range of variables such as gender, second-language proficiency, special educational needs and poverty (via the proxy measure of eligibility for free school meals). However, all these measures have used test outcomes as the basis of calculation and, although statutorily reported, teacher assessments have generally remained in the background.

In early 2007, the government announced a two-year pilot (*Making Good Progress*) which would, among other things, develop a different approach to testing in key

stages 2 and 3. Instead of using a multi-level, end of key stage test to be taken by the full cohort of pupils at exactly the same time, single-level tests in reading, writing and mathematics would be made available on two occasions each year to schools participating in the pilot. Teachers would enter pupils for the relevant test when, through teacher assessment, they judged them to be operating at that level. Thus, at a particular test window, a teacher with a Year 4 class (8-9 year olds), might enter three pupils for a level 4 reading test and five for level 3. This approach had two immediate effects in the pilot schools; it raised the status and profile of teacher assessment by making it the basis of test-entry decisions and it involved teachers from all four years in the key stage instead of simply those who taught pupils in the final year (Year 6).

During 2008-09, it has been decided that the Single-level Test pilot would be confined to key stage 2 and extended to allow the new tests to begin to be used as alternatives to end of key stage full-cohort testing. Thus, from 2009-10, pupils in the pilot schools will take single-level tests instead of the end of key stage tests, initially in mathematics, and have these results reported. This is designed to see what impact a high stakes environment will have on test-entry decisions. If judged to be successful and approved by the new independent assessment regulator, Ofqual, single-level tests will replace end of key stage 2 full-cohort testing.

**Key stage 3 (11-14 years):** As at key stage 2, externally-marked, confidential national tests in reading and writing (including the assessment of Shakespeare), mathematics and science were in place for all 14 year olds (at the end of Year 9) for many years. Following the lead of key stage 2, national targets for improvement in the results were announced by the government and a Secondary National Strategy established to help achieve them. Results were published at school level and value-added measures calculated for pupils' progress in the school between end of key stages 2 and 3 and end of key stages 2 and 4 (using statistics based on results in public examinations at the age of 16). Teacher assessments were statutorily reported and collected nationally but they were given little attention.

In October 2008, the government announced that the piloting of single-level tests at key stage 3 (described above) had been abandoned and that the statutory use of tests at the end of key stage 3 would end with immediate effect. The announcement was accompanied by an exhortation for teachers to use more frequent teacher assessment of pupils throughout the key stage to support their progress and the establishment of an '*Expert Group on Assessment*' which would advise the government on future policy on a number of fronts. Because of the abruptness of the announcement, many schools continued to make use of the national tests which had already been developed for 2009 but the tests were marked by teachers in the school and the results were not reported or collected.

Thus, the current national assessment and moderation arrangements for schools in England can be summarised as follows:

Key Stage	Form of statutory assessment	Moderation	Publication
Early Years Foundation	Teacher Assessment:	Visits to sample of settings each year	National results only

(ages 0-5)	Foundation Stage Profile completed by teachers at the end of final year (Reception)	by Local Authority	
Key Stage 1 (ages 5-7)	Teacher Assessment: Judgements made towards end of final year (Year 2) informed by statutory but flexible use of national tests	Visits to sample of settings each year by Local Authority	National results only
Key Stage 2 (ages 7-11)	Tests and Teacher Assessment: Full cohort tests taken and teacher assessments made at end of final year (Year 6) (ongoing pilot of single-level tests in all year groups which will replace full-cohort end of key stage tests if successful)	None	National and school-level results
Key Stage 3 (ages 11-14)	Teacher Assessment: Judgements made towards end of final year (Year 9)	None	National results only

### **Assessing Pupils' Progress (APP)**

In 2005, the Qualifications and Curriculum Authority (QCA) (now known as the Qualifications and Curriculum Development Agency (QCDA)) began to pilot a new approach to help strengthen teachers' skills in assessing learners' attainment in relation to national standards. Initially in English (reading and writing) at key stage 3, nationally agreed materials were developed to support assessment activity which would value the information already available to teachers, help assessment more directly inform learning, planning and pedagogy and provide the basis for better feedback to pupils and parents. Materials were piloted in a range of schools over two years and first made available via the website of the Secondary National Strategy in 2007. These were followed by resources for mathematics at key stage 3 and then, following further development and piloting, parallel materials were published for reading, writing and mathematics at key stages 1 and 2. Last year, the original key stage 3 materials for reading, writing and mathematics were renewed and joined by others for science and ICT. Piloting of APP for speaking and listening across key stages 1, 2 and 3 and for science in key stages 1 and 2 is now complete and these materials are due for publication in early 2010.

### **APP Materials**

The key elements of the APP approach are sets of criteria known as *Assessment Guidelines* and collections of national exemplification material known as *Standards Files*.

**Assessment Guidelines:** Based on the level descriptions of the National Curriculum's eight-level scales, criteria are set out by assessment focus – critical areas of the subject where learners' progress can be observed. These are not 'tick charts' for use on a regular basis but are designed to be completed for each learner periodically – generally two or three times a year - taking into account the range of available evidence. Completed guidelines reflect the professional judgement of the teacher based on what they have seen and heard and offer a profile of a learner's strengths and areas for development across the subject. It is this profile which can be discussed with the learner and used as the basis for setting targets for improvement. If required, the completed guidelines can also be used to generate an overall level related to national standards. Many schools already gather this kind of information on a termly basis in core subjects in order to track pupils' progress and identify those in need of further support or intervention. The guidelines do not require teachers to undertake specific assessment tasks or activities but encourage them to observe and listen more carefully to learners' behaviour and responses in the course of ongoing teaching and learning and make full use of this information when periodically reviewing progress.

**Standards Files:** In order to support consistent interpretation of criteria and illustrate the range of evidence which can be used to inform teacher assessment, national Standards Files have been developed and published. These differ markedly from previous materials designed to exemplify national standards which have tended to be based around individual pieces of written work from a variety of learners. Standards files are annotated collections of evidence based on the work of individual learners over the course of approximately a term. They can consist of written work, teacher's notes and reflections, photographs, pupil's self-assessments, film-clips and observations from others involved with the learner. A range of files is provided in order to exemplify attainment at a variety of levels and include the relevant APP assessment guideline sheet completed by the teacher. They are intended to be used by schools for the purposes of familiarising teachers with the criteria and the approach used in APP, for standardising interpretation of the criteria and for teachers to use as benchmarks for their own judgements.

### **Implementation and influence of APP**

Alongside the development of APP materials has been an attempt to engage the profession in a different conversation about assessment. This has focused on the need for teachers to be aware of the full assessment repertoire available to them and to be less dependent on tests and tasks. There has been a good deal of investment by schools over recent years in 'assessment for learning' practices and APP has been able to build on and reinforce this investment. QCDA has deliberately avoided use of the terms 'formative' and 'summative' (too often unhelpfully assumed to be in opposition to each other) and has instead referred to 'day-to-day', 'periodic' and 'transitional' as a more helpful set of terms. These terms are now widely used and inform the government's national 'Assessment for Learning Strategy' (DCSF, May 2008). APP is seen to be chiefly concerned with

periodic assessment but the relationship between the three is dynamic and interdependent.

The key features of the three forms are characterised as:

Day-to-day	<ul style="list-style-type: none"> <li>• Learning objectives made explicit and shared with pupils</li> <li>• Peer and self-assessment in use</li> <li>• Pupils engaged in their leaning and given immediate feedback</li> </ul>
Periodic	<ul style="list-style-type: none"> <li>• Broader view of progress across subject for teacher and learner</li> <li>• Use of national standards in the classroom</li> <li>• Improvements to medium-term planning</li> </ul>
Transitional	<ul style="list-style-type: none"> <li>• Formal recognition of pupils' achievements</li> <li>• Reported to parents/carers and next teacher(s)</li> <li>• May use external tasks or tests</li> </ul>

For many teachers who have already placed a greater focus on assessment for learning in their classroom, APP has been the means of making more systematic use of their increased insights into pupils' learning. It has given them a common language and criteria with which to build clear profiles of learners' abilities across a subject; these are reviewed at intervals which are wide enough apart to allow progress to be recognised but also not so late that the insights gained cannot be acted upon to make further improvements.

Although not statutory, APP is now being tried out in a large proportion of schools in England. In many, it is still in the early stages – perhaps it is operating initially in one department of a secondary school or has been started with mathematics across a primary school. The Department for Children, Schools and Families committed itself to making available to schools an additional £50 million for three consecutive years starting in 2008 to help them improve their assessment practice. The focus is on helping schools make a start and recognise the benefits of the approach, not just in providing a more secure basis for teacher assessment but as a means of reviewing curriculum provision and classroom pedagogy. By getting teachers to think harder about the evidence already available to them, many are rethinking the way they question and listen to pupils in group contexts, reconsider the opportunities they give pupils to exercise independence and choice and change how they make use of additional adults in the classroom.

APP has also had a significant influence on the development of single-level tests described on pages 2 and 3 above. Participating pilot schools are required to make termly teacher assessments for all relevant pupils in reading, writing and mathematics and have been advised to use the APP criteria as the basis for these judgements. These in turn inform teachers' decisions about entry for the tests which are then seen as confirmatory of teachers' assessment judgements.

### **Moderation Practice**

Schools are encouraged to look carefully at the Standards Files when they begin exploring APP and to engage in simple standardisation activities but there has been a conscious decision not to place too much emphasis on moderation in the early stages. This was to ensure that teachers, as they began to try out the

approach, did not focus unduly on 'evidence-gathering' because APP does not rely on teachers building extensive portfolios or physically collecting pupils' work; they are encouraged to periodically review the existing evidence of children's learning which arises from good, day-to-day classroom teaching. Experience from the piloting of the approach also suggested that, when first introducing APP, schools were reluctant to drop any current assessment practices and so, in the short term, teachers were faced with 'additional' activity. This appeared to change after about a year when schools were convinced of the APP approach and recognised that it could replace other practice. Typically, the kinds of activity that were dropped were the regular use of optional, teacher-marked national curriculum tests in order to provide pupil tracking information to school management; schools realised that APP could do the same job but with much greater additional benefits to teaching and learning.

Despite this lack of central prioritisation of moderation at this stage, teachers, schools and local authorities are increasingly requesting advice, support and training in this aspect. As they implement the APP approach, they are interested to find out if their judgements are in line with those in other schools and how they can further improve practice.

### **Models of Moderation**

During 2007-08, following the development and piloting of APP in schools, QCDA carried out some exploratory work on different models of moderation. The focus was mainly on different kinds of facilitated meetings for representative teachers from pairs or groups of schools but the initiative also looked briefly at postal and web-based moderation and at visits by external moderators. Typically, Local Authorities organised at least two of three events for participating teachers so that differences could be observed over time.

Teachers were generally enthusiastic about their participation in moderation irrespective of the model used. The opportunity to engage with peers in purposeful, professional dialogue about evidence of learning was seen to be helpful on a number of fronts. Those engaged in supporting the events, either at Local Authority level or from the central team of trainers, also saw significant improvements in the course of the year. They identified several areas where this was evident.

**Accuracy:** over the course of the year, as teachers gained more experience of the processes involved, the accuracy of their judgements in relation to national standards showed general improvement.

**Evidence:** with greater experience of moderation, teachers saw more clearly what kind of evidence was particularly helpful and the nature of contextual information that was provided in good collections. This did not mean that more material was included as time went on but that teachers were better able to discern and make use of the most significant evidence. This was particularly true in two areas where, initially, teachers had often found it hardest to identify appropriate evidence – in Reading and in Using and Applying Mathematics.

**Choice and independence:** there was evidence that a much higher proportion of evidence of fully independent work was included in collections as the year progressed. Where three moderation events took place in the course of the year, the percentage of collections considered to include 'much independent work' was over a third compared to only just over 20% where only two meetings took place.

**Confidence:** participating teachers, particularly those who had been using APP in their classrooms during the pilot phase, appeared to feel increasingly confident in the approach through involvement in moderation. This was seen in the contrast between early events, where some teachers had been reluctant to 'let go' of their collections and preferred to be present when they were discussed, and later meetings where teachers had been much clearer about the benefits of robust discussion and challenge on the basis of the evidence presented.

Moderation activity was also seen as a valuable opportunity to clarify the APP approach and make it more consistently applied across schools and also, for the Local Authority staff, a means of identifying aspects of pedagogy and planning where further professional development was needed locally.

Although some of the models of moderation were tried on a very small scale, the view of the moderators who conducted the work was that there was relatively little difference in the accuracy of outcomes from different models provided certain critical features were in place. These included:

- a sound understanding of what constitutes an appropriate collection for moderation in any given model
- inclusion of a good range of evidence carefully chosen to support specific assessment focus judgements
- inclusion of work where pupils have been able to exercise choice and independence
- sufficient information from the teacher on the context of the evidence included in the collection (particularly where moderation is taking place without the presence of the teacher)
- confidence with the assessment focuses and familiarity with the assessment criteria
- a professional attitude to moderation activity, allowing and promoting challenge to judgements and evidence in order to secure robust, defensible outcomes.

## Conclusions

National curriculum assessment policy and practices in England have changed significantly over the last few years; teacher assessment is now the only form of reported judgement at the ages of 5, 7 and 14 and is becoming more significant in the assessment of 7 to 11 year olds. *Assessing Pupils' Progress* has been designed to strengthen teacher's periodic assessment judgements, build on good assessment for learning practice in day-to-day teaching and help teachers see the integral nature of assessment to their planning and pedagogy. Although not statutory, it provides a national set of criteria and exemplification of standards and is being widely used in schools. There is no requirement for schools to use



moderation when implementing this approach but many are beginning to seek opportunities to discuss and compare their judgements. Evidence from working with a range of APP moderation models suggests that they can lead to increasing accuracy and consistency of judgements, improved practice in terms of curriculum breadth and opportunities for independent work, and greater confidence among teachers both in their understanding of evidence which is particularly significant and in the justification of their assessments.

## References

Qualifications and Curriculum Development Agency (2009) *Assessment*  
<http://www.qcda.gov.uk/13581.aspx>

National Strategies (2009) *Assessing Pupils' Progress*  
<http://www.nationalstrategies.dcsf.gov.uk/primary/assessment/assessingpupilsprogressapp>  
<http://www.nationalstrategies.dcsf.gov.uk/secondary/assessment/assessingpupilsprogressapp>

Department for Children, Schools and Families (2008) *The Assessment for Learning Strategy*, DCSF-00341-2008  
<http://www.teachernet.gov.uk>

Department for Children, Schools and Families (2009) *Report of the Expert Group on Assessment*, 00532-2009DOM-EN  
<http://www.publications.dcsf.gov.uk/eOrderingDownload/Expert-Group-Report.pdf>

Department for Children, Schools and Families (2009) *Making Good Progress Pilot*  
<http://www.teachernet.gov.uk/teachingandlearning/schoolstandards/mgppilot>

Department for Children, Schools and Families (2008) *Evaluation of the Making Good Progress Pilot Interim Report*, DCSF-RR065  
<http://www.dcsf.gov.uk/research/programmeofresearch/projectinformation.cfm?projectid=15619&resultspage=1>