

National Assessment Arrangements 2010/11 Report



November 2011

Ofqual/11/5107

Contents

Introduction: regulating National Assessment arrangements.....	3
Executive summary	6
Early Years Foundation Stage	6
Key Stage 1	7
Key Stage 2	8
Key Stage 3	10
Early Years Foundation Stage.....	11
Background.....	11
Assessment and moderation of EYFS Profile outcomes 2010/11	11
The Tickell Review and its recommendations.....	12
Reports on <i>Cross-phase moderation</i> and <i>Reporting pupils' progress to parents</i> ..	13
Ongoing review activity	14
Key Stage 1	16
Background.....	16
Findings: End of Key Stage 1 assessment cycle 2010/11.....	16
Year 1 Phonics Screening Check	18
Key Stage 2.....	19
Background.....	19
Print and distribution	21
Marking	21
Data and systems	25
Test administration.....	25
Operational security	26

Level setting.....	26
Draft level setting	27
Script scrutiny	28
Final level setting	28
Impact	30
Key Stage 2 test development: 2012 cycle.....	31
Background.....	31
Key Stage 2 English in 2012	34
Optional level 6 tests.....	35
Ongoing review activity	37
Key Stage 3.....	38
Background.....	38
Ongoing review activity	39
Cross-key stage review activity	41
Parental confidence in 3 to 14 assessments.....	41
Implications for future review activity	41

Introduction: regulating National Assessment arrangements

As the regulator of England's examination and assessment system, it is our responsibility to ensure that learners receive fair treatment in the administration of National Assessment arrangements, and that standards are secure and consistent over time.

This report focuses on the activities during 2010/11 of our Early Childhood and National Curriculum Assessment teams which are responsible for reviewing the development and implementation of all statutory assessments in England for 3 to 14-year-olds. It covers our review of the implementation of National Assessment arrangements during the 2010/11 assessment cycle as well as of work undertaken during that period to develop future assessment materials and new arrangements for 2011/12 and beyond.

In February 2011 we published the *Regulatory Framework for National Assessments: National Curriculum and Early Years Foundation Stage*¹ and we now report our annual findings and recommendations for the first time in the context of that document. Two of our statutory objectives inform our work in keeping National Assessment arrangements under review. These are:

- The assessment standards objective: to promote the development and implementation of National Assessment arrangements, which give a reliable indication of achievement, and indicate a consistent level of attainment (including over time) between comparable assessments
- The public confidence objective: to promote public confidence in National Assessment arrangements.

(Regulatory Framework for National Assessments: National Curriculum and Early Years Foundation Stage, paragraph 8, page 6)

In meeting these objectives, we are committed to ensuring that our review activity is proportionate, accountable, consistent, transparent and targeted.

Previous annual reports of this kind, including the one we published on the 2009/10 cycle, focused wholly on the Key Stage 2 statutory tests. These involve the most activity in terms of review because new tests are developed for each cycle and undergo extensive preparatory work. They must be kept confidential until they are taken and are externally marked; this requires a major logistical operation. They also

¹ www.ofqual.gov.uk/files/2011-regulatory-framework-for-national-assessments.pdf

receive considerable attention because outcomes are published at school level and are used for school accountability purposes. The majority of this year's report will also focus on the Key Stage 2 tests but, as our objectives require, we have also looked at and published our findings for all other forms of statutory National Assessment from the Early Years Foundation Stage (EYFS) to the end of Key Stage 3.

The last 12 months have seen an unprecedented amount of proposed change to National Assessment arrangements. A major focus for our activity in 2010/11 has therefore been one of keeping under review proposed changes to future arrangements. As well as keeping under review the statutory National Assessment arrangements which applied in 2010/11, we have also considered the development and piloting of a proposed new National Assessment, the Year 1 Phonics Screening Check, the Review of the Early Years Foundation Stage conducted by Dame Clare Tickell;² the Independent Review of Key Stage 2 testing, assessment and accountability conducted by Lord Bew³ and the introduction of Key Stage 2 optional tests at level 6. We reviewed these proposals and changes in relation to the *Regulatory Framework* in general and to Section 7, paragraphs 65 to 71 in particular. In terms of possible change in the longer term, we have also tried to keep informed about the continuing work of the Review of the National Curriculum in England⁴ insofar as this has within its remit "what, if anything, should replace existing attainment targets and level descriptors to define better the standards of attainment children should reach, and be assessed against, at various points through their education" (Review of the National Curriculum in England: Remit for the review, DfE 2010).

Almost all of the activity to which this report refers was undertaken when the Qualifications and Curriculum Development Agency (QCDA) remained the major responsible body for National Assessment arrangements. The Government had already announced in 2010 its decision to close QCDA and to transfer its assessment responsibilities to an executive agency within the Department for Education (DfE). This change occurred at the beginning of October 2011 when the Standards and Testing Agency (STA) officially came into being. The findings of the

² www.education.gov.uk/tickellreview

³ <https://media.education.gov.uk/MediaFiles/C/C/0/%7BCC021195-3870-40B7-AC0B-66004C329F1F%7DIndependent%20review%20of%20KS2%20testing,%20final%20report.pdf>

⁴ www.education.gov.uk/schools/teachingandlearning/curriculum/b0073043/remit-for-review-of-the-national-curriculum-in-england

report will therefore refer to QCDA's actions but all recommendations are made to the STA and the DfE.

Executive summary

The Office of Qualifications and Examinations Regulation (Ofqual) is the regulator of qualifications, examinations and assessments in England. We are committed to ensuring that standards are maintained and that learners get the results they deserve.

Part of our role under the Apprenticeships, Skills, Children and Learning Act 2009 is to keep under review all aspects of National Assessment arrangements that cover the EYFS Profile and assessment at the end of Key Stages 1, 2 and 3. This report presents the annual findings of our review activities in relation to the *Regulatory Framework for National Assessments* which we finalised after consultation during 2010 and published in February 2011.

Early Years Foundation Stage

Moderation arrangements for the EYFS Profile have become increasingly well established in most local authorities in recent years, in part at least because of national training and support provided by the Qualifications and Curriculum Development Agency (QCDA) and the annual Moderator Registration Programme. There was evidence of a slightly larger number of new local authority colleagues taking responsibility for the management of EYFS Profile moderation than in the previous two years but these colleagues were identified by QCDA and were offered support. There was also evidence, particularly in authorities which made significant reductions in centrally employed staff in the course of the school year, of the scaling back of moderation activities to the minimum required. **However, our overall finding is that arrangements worked smoothly in 2010/11 and that this provides a reasonable basis for confidence that outcomes were accurate and reliable.**

The independent review of the EYFS, led by Dame Clare Tickell, was announced in July 2010. We submitted a response to the call for evidence and subsequently met Dame Clare in September 2010. The final report of the Review was published in March 2011. In July, the Government accepted its main recommendations in full and, between then and September, consulted publicly on a revised draft EYFS. In relation to assessment, the proposals include the introduction of an entirely redesigned EYFS Profile based on “best-fit” judgements in relation to 17 early learning goals. It is proposed that teachers / practitioners must judge, for each child, whether they are at the expected level of development for each goal (“achieving”), above it (“exceeding”) or below it (“emerging”).

In addition, the Review recommended that a new “summary check” be introduced for children between the ages of 2 and 3. This would be undertaken by practitioners and include a judgement of each child’s development and progress in the three prime areas of Personal, social and emotional development, Communication and language

and Physical development. It is not yet clear what the exact status of this check will be.

It is anticipated that changes to the assessment arrangements will apply from 2012/13, with the first statutory use of the new Profile in the summer of 2013.

Our recommendations relate to the proposed new arrangements and focus on the need for the purpose of the assessments to be made explicit and for the intended uses to be stated at the outset. This will help to reduce the risks of different stakeholders ascribing additional purposes to the assessment for which it was not designed and making use of the outcomes in ways which were not intended and which may not be appropriate. We would also expect that adequate piloting of the new criteria and assessment arrangements should take place during 2011/2. This will allow us to undertake appropriate review activity as the assessments are developed and implemented and consider evidence in relation to the five common criteria of validity, reliability, comparability, minimising bias and manageability.

Key Stage 1

End of Key Stage 1 assessment and moderation

Moderation arrangements for end of Key Stage 1 teacher assessment have been in place, largely in their present form, for a number of years. They are well established in most local authorities although there has been limited national training of moderators and, to a large extent, the system has relied on the initiative and collaboration of local authorities in setting up and maintaining regional structures in order to support consistency of judgements across the country.

As was the case with EYFS Profile moderation, there was evidence in 2010/11, particularly in local authorities which made significant reductions in centrally employed staff in the course of the school year, of the scaling back of moderation activities to the minimum required. Because of uncertainty about future changes to National Assessment arrangements and to the National Curriculum itself, some local authorities made temporary arrangements to ensure that statutory minimum requirements were met and were due to review the position at the start of the 2011/2 school year. There is also evidence from the external moderation visits undertaken by consultants working for QCDA that local moderation arrangements do vary considerably, with some local authorities tending to limit the focus of visits to attainment in relation to particular attainment targets and levels.

Despite this level of variation and the degree of uncertainty about the future, our overall finding is that arrangements were satisfactory in 2010/11 and there is a reasonable basis for confidence that outcomes were generally accurate and reliable. Our main recommendation is that the DfE, local authorities and schools should look at ways to strengthen Key Stage 1 moderation, including the

encouragement of greater involvement of teachers in Year 3, in order to ensure that it leads to accurate and robust outcomes which provide as reliable a basis as possible for calculating pupils' progress by the end of Key Stage 2.

Year 1 Phonics Screening Check

The Government signalled its intention to introduce a new form of National Assessment in the White Paper which it published in November 2010. This was to take the form of a Phonics Screening Check designed to be used as a full-cohort assessment by all schools attended by Year 1 pupils. Plans were announced to pilot the new assessment in the summer of 2011 with a view to introducing it nationally in the summer of 2012.

We held some discussions at an early stage with the DfE, which undertook the development of the Phonics Screening Check, to clarify the nature and scope of the assessment. We also set out the range of evidence which we would need to see in order for this proposed new National Assessment to meet the five common criteria described in the *Regulatory Framework for National Assessments* (Ofqual, 2011), namely validity, reliability, comparability, minimising bias and manageability. This was reflected in the publication by the DfE of a detailed Framework document for the Phonics Screening Check in March, 2011.⁵

We had good access to the arrangements for piloting the Check in the summer of 2011. We were provided with initial statistical and qualitative evidence from the pilot in October 2011 and have provided interim feedback. We are awaiting the full analysis of evidence from the pilot and the final outcome of the standard-setting process early in 2012.

Given that the Phonics Screening Check will be the first new form of National Assessment for a number of years, our recommendation is that the DfE and STA draw up clear plans to monitor its early implementation both to ensure that it is fulfilling its declared purpose and to consider its impact on school practice and behaviour.

Key Stage 2

Key Stage 2 arrangements 2011

Between 2010 and 2011, there were no significant changes to the requirements for pupils to undertake statutory tests in reading, writing and mathematics and to be teacher-assessed in English, mathematics and science. Unlike the previous year, when approximately 25 per cent of schools in England boycotted the statutory tests, all schools administered them in May 2011. At the same time, the Bew Panel, set up

⁵ www.education.gov.uk/schools/teachingandlearning/pedagogy/a0077650/phonicsframework

by the Government to look at testing, assessment and accountability at Key Stage 2, was preparing to publish its final report and recommendations.

Overall, the assessment arrangements for Key Stage 2 statutory assessment in 2011 were successful and there was no point when we needed to consider exercising our duty to notify the Secretary of State for Education that there was likely to be a significant failing. There is a secure basis for confidence that outcomes were generally accurate and reliable and comparable with those from previous years. Marking was completed in good time, results were provided for schools as planned and almost all teacher assessment data was collected before the end of the summer term. This is particularly noteworthy in that QCDA was overseeing the delivery of the assessment arrangements at the same time as it was preparing to reconfigure and transfer its assessment operations into the new executive agency of the Department for Education, the Standards and Testing Agency.

We make a number of detailed recommendations and suggestions aimed at further improving the quality of delivery in the future. We also seek to ensure access to relevant information and documentation and will be working with the STA and the DfE to put appropriate arrangements in place. We recognise, however, that significant changes to the National Assessment arrangements will be made both in 2012 and again in 2013 and our review activity in 2010/11 included a strong focus on the implications of these changes.

Developments for 2012 and beyond

We had observer status on the Bew Review, set up by the Government in the autumn of 2010 to enquire into testing, assessment and accountability at the end of Key Stage 2. We were therefore aware of the potential implications for National Assessment arrangements of the Review's recommendations which were published in June of this year and accepted in the Government's response in July. We fully understand the difficulties faced by the DfE and STA in balancing rapid implementation of the proposals with giving the necessary time to develop assessment arrangements and tests which provide valid, reliable and fair outcomes for learners. The interim arrangements for 2012 have become clear over the last few months; they involve a number of changes including:

- teacher assessment being the only reported assessment for English writing but informed by the statutory use of a nationally provided test
- the introduction of some form of external moderation for the teacher assessment of writing
- the use of a national sampling writing test in approximately ten per cent of schools in order to monitor national standards

- the availability of optional level 6 tests which, for reading and mathematics, will be externally marked and the outcomes of which can be used for school accountability purposes.

These are complex changes to be introduced in a single year and there is a risk that schools and local authorities will not grasp the full implications or will do so at a point when there is limited time to plan effectively for them. **We acknowledge that the Standards and Testing Agency has sought to minimise this risk by putting in place strategies to communicate with stakeholders but this needs to be a continuing focus for its work over the coming months and we will keep this under active review.** We make a number of recommendations relating to developments for 2012, both in relation to English writing and to the optional level 6 tests for reading and mathematics.

Key Stage 3

There are no statutory moderation arrangements in place at Key Stage 3 and so it is hard for us to make any definitive statement about our confidence in the robustness and accuracy of statutory assessment outcomes. We noted an increased effort by the QCDA to collect Key Stage 3 teacher assessment data. As a result of repeated contact by the QCDA with schools and local authorities during the late summer of 2011, over 99 per cent of data had been submitted by the start of the autumn term, a higher percentage than in previous years.

We have been undertaking some evidence gathering from schools, local authorities and other stakeholders about current practice in statutory Key Stage 3 assessment, particularly since the end of the requirement, at the end of 2008, for pupils to undertake externally marked tests in English, mathematics and science. It appears that this change, along with the ending of requirements for schools and local authorities to set targets for performance at the end Key Stage 3 and of the publication of Key Stage 3 performance tables, has meant that statutory Key Stage 3 assessments have become a much lower priority for schools with the exception of the minority of schools where Year 9 is the final year and pupils move to other institutions at the start of Year 10.

We have not made any specific recommendations in relation to Key Stage 3 assessment arrangements but will complete our evidence gathering exercise and consider what review activity would be appropriate during the remainder of this school year.

Early Years Foundation Stage

Background

The current EYFS Framework has been in place since September 2008 and this was therefore the third full year of use for the EYFS Profile which practitioners must complete for each child in their final term of the Reception year, before the transition to Year 1. The requirements for practitioners are set out in the *Statutory Framework for the Early Years Foundation Stage*⁶ (DCSF, May 2008) and further guidance is provided in the *Early Years Foundation Stage Profile Handbook* (QCDA, 2008). The stated purpose of the Profile assessment is, “ascertaining what [children] have achieved in relation to the early learning goals” (Childcare Act 2006, paragraph 41, 2c).

Assessment and moderation of EYFS Profile outcomes 2010/11

QCDA maintained its support for Profile moderation during 2010/11 by collecting and analysing local authority plans, providing support where weaknesses or succession issues were identified and organising external moderator visits to a sample of local authorities. It also ran seven regional events as part of its Moderator Registration Programme and, as in previous years, invited all local authorities to nominate two colleagues to attend. This was the third full year that these assessed two-day events have been run and a further 179 moderators participated this year meaning that almost 700 practising local authority EYFS Profile moderators have now been fully accredited through successful participation.

Findings

We have received monthly updates on QCDA activity relating to EYFS Profile assessment at the QCDA Programme Board throughout 2010/11 and have had opportunities to raise specific questions at our quarterly accountability meetings with senior officers from QCDA. We have also been invited to attend as observers the EYFS Profile delivery meeting, hosted by DfE, where practical issues of implementation were addressed. Such meetings stopped at the end of the 2010/11 cycle but have now restarted under the new agency arrangements.

We attended meetings held at QCDA to evaluate local authority plans for moderation and to debrief the external moderators who had conducted visits to a sample of local authorities to see moderation in practice and discuss local arrangements. We observed one of the Moderator Registration events and discussed current local

6

www.education.gov.uk/childrenandyoungpeople/earlylearningandchildcare/delivery/education/a0068102/early-years-foundation-stage-eyfs

moderation practice both with members of our Early Years Advisory Group and on visits to meet colleagues working in the 12 representative settings and schools with which we retain ongoing links.

Moderation arrangements for the EYFS Profile have become increasingly well established in most local authorities in recent years, in part at least because of national training and support provided by QCDA and the annual Moderator Registration Programme. There was evidence of a slightly larger number of new colleagues taking responsibility for the management of EYFS profile moderation than in the previous two years but these colleagues were identified by QCDA and were offered support. There was also evidence, particularly in authorities which made significant reductions in centrally employed staff in the course of the school year, of the scaling back of moderation activities to the minimum required. However, our overall finding is that arrangements worked smoothly in 2010/11 and that this provides a reasonable basis for confidence that outcomes were accurate and reliable.

Although local authorities appear to have met statutory requirements in 2010/11, there was considerable uncertainty expressed about the level of support which would be available in 2011/12. For example, one local authority moderation manager said, "With the restructuring of the LA I am not sure of my future role. This means I can't plan for next year because of this. I am not sure who will be available to deliver moderation next year."

The Tickell Review and its recommendations

The independent review of the EYFS, led by Dame Clare Tickell, was announced in July 2010. We submitted a response to the call for evidence and subsequently met Dame Clare in September 2010. The final report of the Review was published in March 2011 and the Government's response in July accepted the main recommendations in full. Between July and September 2011, the Government consulted on a revised draft EYFS. In relation to assessment, the proposals include the introduction of an entirely redesigned EYFS Profile based on "best-fit" judgements in relation to 17 early learning goals. It is proposed that teachers / practitioners must judge, for each child, whether they are at the expected level of development for each goal ("achieving"), above it ("exceeding") or below it ("emerging").

In addition, the Review recommended that a new "summary check" be introduced for children between the ages of 2 and 3. This would be undertaken by practitioners and include a judgement of each child's development and progress in the three "prime" areas of Personal, social and emotional development, Communication and language and Physical development. It is not yet clear what the exact status of this check will be.

The Government's public consultation on the proposed changes ended on 30th September 2011 and it is anticipated that changes to the assessment arrangements will apply from 2012/13, with the first statutory use of the new Profile in the summer of 2013. We have made it clear that, in relation to the proposed new Profile, it is crucial that the purpose of the assessment is made explicit and that the intended uses are stated at the outset. This will help to reduce the risks of different stakeholders ascribing additional purposes to the assessment for which it was not designed and making use of the outcomes in ways which were not intended and which may not be appropriate. We would also expect that adequate piloting of the new criteria and assessment arrangements should take place during 2011/2. This will allow us to undertake appropriate review activity and consider evidence in relation to the five common criteria of validity, reliability, comparability, minimising bias and manageability.

Recommendations / areas for improvement

We recommend that the DfE, in finalising the statutory arrangements for assessment in the revised EYFS, clearly sets out:

- the purpose of the assessment
- the uses for which the assessment outcomes are intended
- the legal requirements for both the proposed assessment at 24 to 36 months and the assessment at the end of the EYFS for practitioners, for settings and schools and for local authorities
- who will be responsible for both developing and implementing the assessments in such a way that they meet the common criteria of the *Regulatory Framework*
- who will be responsible for monitoring and quality assuring the assessments in order to ensure that reported outcomes are accurate and robust.

This information would form the basis of our regulatory work in this area and bring the new EYFS assessment arrangements in line with international standards of assessment development.

Reports on *Cross-phase moderation and Reporting pupils' progress to parents*

We undertook two pieces of background review activity which led to the publication of two reports on our website: *Cross-phase moderation at the end of the Early Years*

Foundation Stage,⁷ and *Reporting pupils' progress to parents in the Early Years Foundation Stage*.⁸ Both reports were also shared in draft form with Dame Clare Tickell's Review early in 2011.

Cross-phase moderation at the end of the Early Years Foundation Stage: This report focused on the nature and extent of cross-phase moderation between EYFS practitioners in Reception and Year 1 teachers in advance of pupils' transition to Key Stage 1. Although there is no statutory requirement for such moderation, it appears to have become more widespread in recent years and qualitative evidence from the structured interviews carried out as part of the study suggested that it had led to both increased confidence in the accuracy and robustness of EYFS Profile judgements and to better use of Profile information by receiving teachers in Year 1. We noted that the subsequent report of the Bew Review (into Key Stage 2 testing, assessment and accountability) also called for greater use by schools of cross-phase moderation activity in order to improve transition between phases by helping to ensure that pupil-level data is robust and that receiving teachers feel confident in making wide use of it to understand their new intake.⁹

Reporting pupils' progress to parents in the Early Years Foundation Stage: This report, based on interviews and an electronic survey of the views of practitioners and parents, considered the range of ways in which settings and schools provide information about children's progress, including the outcomes of the EYFS Profile, to parents. The report concluded that, while parents generally felt confident about the information they received about their children's progress during and at the end of the EYFS, this was provided in a wide variety of forms and in some cases did not make explicit reference to the early learning goals.

Ongoing review activity

Reviewing local authority capacity: The DfE is responsible for ensuring that local authorities make provision to ensure the accuracy and consistency of the assessments made by Early Years providers in their area. Evidence from external moderation undertaken by QCDA in 2010/11 indicated that many local authorities

⁷ www.ofqual.gov.uk/files/11-04-12-Cross-phase-Moderation-at-the-End-of-the-Early-Years-Foundation-Stage.pdf

⁸ www.ofqual.gov.uk/files/11-04-12-Reporting-Pupils-Progress-to-Parents-in-the-Early-Years-Foundation-Stage.pdf

⁹ See "Transition from Key Stage 1 to Key Stage 2", *Independent Review of Key Stage 2 testing, assessment and accountability: Final report*, Lord Bew, page 69 (see footnote 3 above for a link to the Bew Report).

have reduced moderation activity for the EYFS Profile to a minimum and there were anxieties expressed about whether these arrangements could be maintained in 2011/12. As far as evidence is available, we will continue to keep this situation under review. For 2011/12, this will include the information, training and guidance materials which will be developed and made available to support the introduction of the new assessment arrangements for the EYFS which are due to come into force in September 2012.

Evaluation of the Moderator Registration Programme: The purpose of this project was to gather participants' views on the impact on standards and assessment practice of the national EYFS Profile Moderator Registration Programme which has been run annually by QCDA to accredit practising local authority Profile moderators in each of the last three years. We undertook 20 semi-structured interviews with people who had completed the programme in 2009 or 2010 and used a large-scale questionnaire to gain further quantitative evidence. We published a report¹⁰ in November 2011 with our findings and recommendations to the DfE.

Early Years Advisory Group and retainer settings: Our Early Years Advisory Group met in May and October 2011. The group is made up of representatives from a range of Early Years stakeholders including practitioners, headteachers, local authorities and training providers. The group helped to contribute towards evidence for the Tickell Review and for monitoring local authority capacity.

We have ongoing links with a sample of 12 Early Years settings from around England. We visit these settings twice a year in May and October and use the visits to collect qualitative evidence which is fed into team projects. For the May visits we collected evidence for the Tickell Review, local authority capacity and moderation practice and from the visits in October we gathered reactions to the proposed check of progress for 2- to 3-year-olds and the Year 1 Phonics Screening Check and also explored influences on parental confidence in assessment outcomes.

¹⁰ www.ofqual.gov.uk/files/11-11-10-Moderator-Registration-Programme-Evaluative-Report.pdf

Key Stage 1

Background

The current approach to statutory assessment at the end of Key Stage 1 has been in place since 2005 when the outcomes of statutory tests and tasks, instead of being separately reported as had been the case until then, were used as part of the evidence on which teacher assessment judgements were made at the end of Year 2. Teacher assessments of pupils' attainment in English (reading, writing and speaking and listening), mathematics and science are the only outcomes now reported to parents and submitted to the DfE. These outcomes also form the basis of school accountability measures in relation to rates of progress made between Key Stage 1 and Key Stage 2; these appear in the annual published performance tables and are considered by Ofsted school inspectors as part of the evidence for the effectiveness of the school.

Local authorities are required to externally moderate assessment arrangements and judgements in all their schools over a four-year cycle. QCDA has had a remit to oversee these local moderation arrangements. It has collected and evaluated local authority moderation plans and has also commissioned external moderators to visit a sample of authorities each year. These visits generally include an opportunity to accompany a local moderator on a school moderation visit.

The use of tests and tasks remains statutory but these are used flexibly at the discretion of the school and the outcomes are then taken into account when teachers are making their end of Key Stage assessments. New tests and tasks are not developed each year, as was the case until 2005 and as remains the case at the end of Key Stage 2, and so those from previous years are re-used. Schools have a constrained choice each year on the tests and tasks that can be used.

Key Stage 1 assessment is the only point in statutory National Curriculum assessment where teachers are asked to make distinctions between pupils' performance within a level. At level 2, where the large majority of 7-year-olds are judged to be operating by the end of the key stage, teachers have to make a "best-fit" judgement in relation to 2c (just into level 2), 2b (securely within level 2) and 2a (strong performance at level 2 and meeting some aspects of level 3).

Findings: End of Key Stage 1 assessment cycle 2010/11

We have received monthly updates on QCDA activity relating to end of Key Stage 1 assessment at the QCDA Programme Board throughout 2010/11 and have also had opportunities to raise specific questions at our quarterly accountability meetings with senior officers from QCDA. We attended meetings at the DfE where QCDA provided updates on their moderation activity and also observed the debriefing meeting for

QCDA's external moderators following their visits to local authorities in the summer term of 2011.

Moderation arrangements for end of Key Stage 1 teacher assessment have been in place, largely in their present form, for a number of years. They are well established in most local authorities although there has been limited national training of moderators and, to a large extent, the system has relied on the initiative and collaboration of local authorities in setting up and maintaining regional structures in order to support consistency of judgements across the country.

As was the case with EYFS Profile moderation, there was evidence in 2010/11, particularly in local authorities which made significant reductions in centrally employed staff in the course of the school year, of the scaling back of moderation activities to the minimum required. Because of uncertainty about future changes to National Assessment arrangements and to the National Curriculum itself, some local authorities made temporary arrangements to ensure that statutory minimum requirements were met and were due to review the position at the start of the 2011/2 school year. There is also evidence from the external moderation visits undertaken by consultants working for QCDA that local moderation arrangements do vary considerably, with some local authorities tending to limit the focus of visits to attainment in relation to particular attainment targets and levels.

Despite this level of variation and the degree of uncertainty about the future, our overall finding is that arrangements were satisfactory in 2010/11 and there is a reasonable basis for confidence that outcomes were generally accurate and reliable. However, schools and local authorities continued to highlight concerns about the perverse incentive operating in primary schools attended by pupils in both Key Stage 1 and Key Stage 2; this may lead to pressure for judgements about attainment of pupils at the end of Year 2 to err on the side of caution given that this might subsequently help to improve value-added scores and progress measures when the same cohort reaches the end of Year 6. The issue was raised in evidence submitted to the Bew Panel and will continue to be an area of focus for our review activity.

Recommendations / areas for improvement

In the light of the strongly argued view from the Bew Panel that pupils' progress should be given at least as much consideration as their attainment at the end of Key Stage 2, we recommend that the DfE and local authorities should look at ways to strengthen Key Stage 1 moderation, including the encouragement of greater involvement of teachers in Year 3, in order to ensure that it leads to accurate and robust outcomes which provide as reliable a basis as possible for calculating pupils' progress by the end of Key Stage 2.

Year 1 Phonics Screening Check

The Government signalled its intention to introduce a new form of National Assessment in the White Paper which it published in November 2010. This was to take the form of a Phonics Screening Check designed to be used as a full-cohort assessment by all schools attended by Year 1 pupils. Plans were announced to pilot the new assessment in the summer of 2011 with a view to introducing it nationally in the summer of 2012.

Findings: Year 1 Phonics Screening Check

We held some discussions at an early stage with the DfE, which undertook the development of the Phonics Screening Check, to clarify the nature and scope of the assessment. We also set out the range of evidence which we would need to see in order for this proposed new National Assessment to meet the five common criteria described in the *Regulatory Framework for National Assessments* (Ofqual, 2011), namely validity, reliability, comparability, minimising bias and manageability. This was reflected in the publication by the DfE of a detailed Framework document for the Phonics Screening Check in March 2011.¹¹

We attended meetings of the Phonics Screening Check Steering Group at the DfE throughout 2011 as well as the training offered in May and June to Year 1 teachers in schools which were part of the pilot of the Check in the summer term. We also observed the initial meeting in September 2011 where the standard for this new assessment was defined and the level for the pass mark provisionally set. We were provided with initial statistical evidence from the pilot in October 2011 and have scrutinised other available evidence including the external evaluation of the pilot conducted on behalf of the DfE by colleagues from Sheffield Hallam University. We have provided interim feedback and are awaiting the full analysis of evidence from the pilot and the final outcome of the standard-setting process early in 2012.

Recommendations / areas for improvement

Given that the Phonics Screening Check will be the first new form of National Assessment for a number of years, we recommend that the DfE and STA draw up clear plans to monitor its early implementation both to ensure that it is fulfilling its declared purpose and to consider its impact on school practice and behaviour.

¹¹ www.education.gov.uk/schools/teachingandlearning/pedagogy/a0077650/phonicsframework

Key Stage 2

Background

Many aspects of the statutory arrangements for end of Key Stage 2 assessment have been in place since 1995. Outcomes of statutory tests and teacher assessment in English and mathematics have been reported separately to parents and gathered centrally for publication by the DfE over this period. They are used for the purpose of school accountability both in terms of overall attainment at the end of the primary phase and also, in more recent years, in relation to the progress of pupils from the end of Key Stage 1 to the end of Key Stage 2. These measures are produced in the annual performance tables compiled by the DfE and are considered carefully by Ofsted school inspectors.

For English, mathematics and science, the end of Key Stage 2 arrangements for 2010/11 were largely unchanged from the 2009/10 cycle when the full-cohort science test had been replaced by a national sampling test undertaken by Year 6 pupils in a random sample of five per cent of schools nationally. In contrast to English and mathematics statutory assessments, the outcomes of this science sample test are intended to be used only for monitoring national standards and not for school accountability. The science sample test was therefore in its second full year but some changes were made to its administration including the return of individual pupils' results to the sample schools and the reporting of these to parents.

For the first time, schools could access level 6 tests in reading, writing and mathematics from the QCDA website. Their use was entirely optional; outcomes were not collected and there was no requirement to share outcomes beyond the school.

There were no significant changes to the requirements for pupils to undertake statutory tests in reading, writing and mathematics and to be teacher-assessed in English, mathematics and science. Unlike the previous year, when approximately 25 per cent of schools in England boycotted the statutory tests, all schools administered them in May 2011. At the same time, the Bew Panel, set up by the Government to look at testing, assessment and accountability at Key Stage 2, was preparing to publish its final report and recommendations.

For 2011, QCDA was the responsible body appointed by the Secretary of State for Education to develop and implement the end of Key Stage 2 tests in mathematics and English, and the end of Key Stage 2 science national sample test. Statutory teacher assessment judgements, which have been included in the published performance tables since 2010, do not have to be submitted by schools until they have received their test results for English and mathematics and have not been subject to any formal moderation arrangements.

Findings: 2011 test delivery overview

In line with previous practice in 2009 and 2010, QCDA has continued to review and quality assure its own progress and that of its suppliers towards the timely delivery of the statutory tests. This practice has provided us with clear evidence of effective planning informed by a coherent “lessons learned” process for each of the key delivery strands of print and distribution, marking, data and systems, test administration and operational security. The approach adopted by QCDA in terms of programme and project management has been key to ensuring that delivery remained focused. Overall, the expectations of paragraphs 90 and 91 of the *Regulatory Framework* which set out the principles for implementation of each area of work have been met.

Overall, the assessment arrangements for Key Stage 2 statutory assessment in 2011 were successful and there was no point when we needed to consider exercising our duty to notify the Secretary of State for Education that there was likely to be a significant failing. There is a secure basis for confidence that outcomes were generally accurate and reliable and comparable with those from previous years. Marking was completed in good time, results were provided for schools as planned and almost all teacher assessment data was collected before the end of the summer term. This is particularly noteworthy in that QCDA was overseeing the delivery of the assessment arrangements at the same time that it was preparing to reconfigure and transfer its assessment operations into the new executive agency of the Department for Education, the Standards and Testing Agency.

A particularly successful aspect of QCDA’s delivery in 2011 was its identification and management of issues. When issues likely to impact on delivery arose, QCDA alerted us to them transparently and in a timely manner, and managed them proactively and appropriately. For example, QCDA’s own quality assurance procedures led to the early identification and resolution of printing issues which could have had a significant impact on pupils’ experience of the end of Key Stage 2 tests.

Although QCDA reported escalated risks as part of its regular update on overall programme delivery and invited us to meetings of its Risk Committee, our access to its full risk register was limited. Since our approach to keeping assessment arrangements under review is risk-based, as set out in paragraphs 45 to 51 of the *Regulatory Framework*, we need to retain full visibility of the risks relating to those arrangements. During 2010/11, QCDA’s reporting of selected risks has not fully provided us with the information required for us to perform our duties independently, efficiently and effectively as set out in paragraphs 19 to 21 of the *Regulatory Framework* but we are working with STA to resolve the necessary practical arrangements to make this possible.

Print and distribution

The QCDA print and distribution function is responsible for ensuring that all pupils receive quality materials for Key Stage 1 and 2 tests which are accurate and that the movement of test materials remains secure at all times throughout the production, testing and marking processes.

Findings

For the 2011 delivery cycle, we kept this function under review primarily by following weekly delivery updates during the critical periods.

Early in 2011, there was an issue in accounting for end of Key Stage 1 materials following printing. This resulted in the process for printing end of Key Stage 2 materials being revised to ensure that all materials were accounted for and security was not compromised.

Due to Easter and the early May bank holiday falling close together in 2011, there was significantly greater variation in the times when schools closed for holidays. Although this presented a considerable planning issue for QCDA, we noted that the complex delivery schedule for test materials was followed with no resulting impact on schools.

With a large scale multiple distribution operation involving some 16,000 schools and approximately 3,500 markers, accounting for all materials is a significant challenge. To a large extent, the process was managed successfully with few materials going missing. Where issues did arise, these were usually the result of human error with processes and protocols not being adhered to by contracted personnel.

As we have already noted there were some problems with print quality this year. Although these were remedied in good time, we are concerned that the issues were usually picked up by QCDA staff and not by the supplier responsible.

Recommendations / areas for improvement

STA should continue to use lessons learned from previous activity to inform future planning in order to ensure that quality and timely delivery are maintained, especially with regard to print quality and security, and to reduce the opportunity for human error to impact on processes.

Marking

The QCDA marking function is responsible for the management of the test operations agency in its delivery of end of Key Stage 2 tests for mathematics and English. In addition, QCDA directly managed the delivery of the end of Key Stage 2 science national sample test.

For the 2011 delivery cycle we kept this function under review by following the weekly updates and by attendance at selected meetings. Selection of meetings to attend was based on our understanding of key risk areas – the training of the senior markers during Meeting 4 and the training of markers during Meeting 8. Observation of these training meetings provided us with a snapshot of attitudes and practices. We also attended the regular operational meetings on a targeted basis and considered the notes of the meetings that we did not attend. We ensured that we were familiar with the Reviews Process document and attended the Review Briefing session, at which key messages are relayed to the Marking Review Panel.

Findings: QCDA marking function

A key success of the marking function in its delivery of the 2011 tests was the timely manner in which it completed its activities. Processes introduced in 2010 to improve the development of marker training and quality assurance materials were used again in 2011.

The recruitment of markers for all end of Key Stage 2 tests was completed well in advance of the planned deadline and markers who were unable to fulfil their commitment at late notice were replaced before the commencement of marker training. As in 2010, QCDA was not able to supply us with a profile of the marker workforce in terms of age, ethnicity, previous experience and so on. Although individual markers are not obliged to provide all of this information, it means that a potential source of important information to help with future recruitment and succession planning remains untapped.

We found that all the marker training venues where we attended events were fit for purpose and the administration of the marker training meetings for English, mathematics and science were well managed and efficient. In particular, good security measures were in place and actively followed throughout by both administrative staff and meeting leaders, with the exception of one venue, where no sign-in of confidential materials provided for the use of the Ofqual observer was requested.

Recommendations / areas for improvement

In our 2010 report, we recognised the effectiveness of the online marking systems used for the Key Stage 2 science sampling test in supporting reliability of outcome. This was replicated in 2011. We would recommend the appropriate expansion of these systems to other aspects of the Key Stage 2 testing arrangements as future circumstances allow.

With one exception, we noted the consistent application of security protocols at all the venues we visited. Given the confidentiality of the test materials, we recommend

that the importance of security and confidentiality continues to be reinforced throughout the marking process.

Findings: English and mathematics marking process

We attended meetings towards the beginning of the marker training cascade for English and mathematics where senior markers are trained and the final meetings where the majority of markers are trained. For both subjects, we found that the training materials were appropriate for purpose and used consistently throughout; meeting leaders made good use of the commentary provided to support training with frequent reference to mark schemes; advice on achieving marking consistency was a frequently repeated theme. Observers noted significant variation, however, in the approach adopted by marker team leaders, with some reading directly from the prepared script, whilst others seemed to use the script as a framework and adjusted their delivery in response to team members' needs in order to facilitate and maximise markers' active engagement with the material.

Whilst most marker teams seemed to be well balanced with the majority being experienced markers, we noted one team which consisted of a majority of new markers and which was led by a first-year team leader.

For English, as in previous years, our observation indicated that meeting leaders found it hard to cover adequately two very different mark schemes – for reading and writing – in the time available. This put pressure on team leaders to follow closely the agenda timings and, as a result, there was limited opportunity to develop sustained discussion and active engagement with the tests and training materials. In particular, the afternoon sessions on training to mark writing did not provide as much evidence of marker engagement – through discussion and questioning – as those dealing with reading which took place during the morning session.

The quality assurance process designed to monitor each marker's application of the mark scheme at the agreed standard during the marking process remained the same as that used in 2010. The introduction of the online collection of marks for electronic return by markers was successful; it allowed increased visibility of marker activity and more rapid intervention by team leaders where necessary.

As in 2010, each review panel member was re-standardised at home and trained to apply review procedures before admittance to the panel. We attended the marker review panels for both mathematics and English (reading and writing). Both review panel briefings were consistent in covering key messages for panellists, though there were slight differences in the use of language between the subject panels with the term "re-mark" being used in place of "review" at the mathematics briefing. We also noted the encouragement for both sets of panellists to conduct reviews at pace. There were clear systems and procedures in place for logging scripts and recording outcomes administered by test operations agency staff. This allowed panellists to

focus on their task which was to review each script to ensure that the original marker had applied the mark scheme at the agreed standard. We noted the security of the reviews process was robust.

It was clear that care was taken to avoid panellists reviewing scripts originally marked within their own team. We also noted that, while review marking is predominantly conducted by a level above the original marker, this is not possible in the case of the marking programme leaders (the highest rank of marker), who, in such a system, will inevitably review their own original script. This is mitigated to some extent by the deputy or assistant marking programme leader also reviewing these scripts.

Recommendations / areas for improvement

In 2012, the arrangements for end of Key Stage 2 assessment of English writing will be different and this provides an opportunity for change. We recommend that the training of markers for reading and writing addresses the issues relating to time allocation noted by us in 2009 and 2010, and again in 2011.

Whilst it is obviously important to produce a script with key messages and training points for all trainers to follow in order to ensure consistency, we would recommend that trainers should also be encouraged to be responsive to the particular needs of their own teams so that markers also have the opportunity to engage actively with the mark scheme and training materials through discussion and questioning. This is particularly important for English where the training day is currently longer than for mathematics.

We would recommend that consideration is always given to the make-up of each team in order to ensure that team leaders have the opportunity to meet the demands of the training schedule and to ensure that all their markers are adequately prepared for the next stage of the marking process.

In our 2010 report, we noted that “It was ... possible for marking programme leaders to review requests submitted by schools against their own original marking.” This remained a possibility in 2011. We recommend that the review marking process for 2012 is revised to ensure that this possibility is removed from the system. Likewise, the system of home-based restandardisation should be reconsidered in order to ensure transparency in the review marking process. Additionally, we would recommend that as well as stressing the need for efficiency during the review marking process to complete the process in a timely manner, panellists should equally have their attention drawn to the need to ensure quality and fairness to pupils throughout the process.

Findings: science marking process

For science, we attended two meetings; the first for training team leaders to mark and the second for training markers to mark. The training materials and associated

activities allowed markers to engage with the tests during training and markers appeared to be confident about their understanding of the science mark schemes. Overall, the online marking of science has shown itself to be an efficient, quality-driven process.

Data and systems

The QCDA data and systems function has two key responsibilities. It ensures that each system used to deliver assessment arrangements for Key Stages 1, 2 and 3 is user-friendly, robust and fit for purpose. It also ensures that data generated by the assessment arrangements for Key Stages 2 and 3 is collected, quality assured and disseminated (arrangements for the collection of data at Key Stage 1 and from the EYFS Profile have been overseen directly by DfE).

Findings

For the 2011 delivery cycle we did not attend any specific meetings relating to data and systems. We kept this function under review primarily by closely following weekly delivery updates and clarifying points through the QCDA National Curriculum Assessments Programme Board and our quarterly accountability meetings with QCDA.

Throughout the delivery period each system was procured, developed and tested in a timely manner and data was delivered in line with requirements. However, the platform for delivering these systems to schools, NCA Tools online, did not always function as anticipated and this created access difficulties for a number of schools.

QCDA also identified that aspects of its 2011 arrangements for schools which applied for access arrangements for individual pupils were non-compliant with data protection legislation. This issue is being addressed by STA for the 2012 arrangements.

Test administration

Test administration includes four separate functions which work together to support schools and local authorities. These are: communications, stakeholder relations, test administration and the National Curriculum helpline. For the 2011 test cycle, it also included a small local authority support team.

Findings

For the 2011 delivery cycle we kept this function under review primarily by following weekly delivery updates, by attending reference group meetings and by reviewing all communications to stakeholders.

Overall we found the communications generated by the test administration function to be accurate and accessible to schools and other stakeholders. The help desk was

well briefed, functioned effectively even when capacity was fully stretched during peak delivery times, and was responsive to issues identified by stakeholders.

Operational security

This function is responsible for security across all aspects of delivery relating to assessment arrangements.

Findings

For the 2011 delivery cycle we kept this function under review primarily by following weekly delivery updates, and also by considering security as we attended a range of meetings throughout the process. Where specific security issues were identified we have raised these in previous sections of this report.

Overall we found that the security function was effective in anticipating issues relating to the delivery of assessment arrangements with clear procedures and protocols in place for managing issues. When security issues did occur, we were notified in an appropriate and timely manner and mitigations were put in place to reduce the impact of these issues. However, we were concerned that security resources might have been insufficient had issues escalated significantly.

Recommendations / areas for improvement

Contingency plans should make provision for deployment of additional security resources if issues escalate in future delivery of assessment arrangements.

Level setting

Level setting is the main process for maintaining standards in National Curriculum tests. The level setting process involves consideration of a variety of quantitative and qualitative data to set threshold marks for each National Curriculum level covered by the test to ensure that the established standard is maintained year on year. This means that, irrespective of the year the test is taken, the standard of performance required to obtain each National Curriculum level should be the same. The threshold mark may change from year to year to reflect the relative difficulty of the test, but the actual standard required to obtain, for example, a level 4 in mathematics, remains the same.

Findings

We considered the level setting process in 2011 in relation to the expectations of the *Regulatory Framework* which states that, “the responsible body or bodies [are expected to] provide exemplification of standards for each National Assessment they develop and to ensure that these standards are maintained” (paragraph 82). From our observations and review activities, we are satisfied that the level setting process for the Key Stage 2 National Curriculum tests in 2011 ensured that standards were maintained.

Draft level setting

The test development agencies pre-test each National Curriculum test a year before it is used with Year 6 children. This allows them to propose draft thresholds for each level using a variety of statistical methods. The evidence in support of these threshold marks is presented at the draft level setting meeting by the test development agency (TDA). Each subject has its own draft level setting meeting and is attended by the TDA, QCDA staff and by an independent adviser appointed by QCDA. We also attend these meetings as observers.

The participants at these meetings evaluate the evidence presented by the TDA and agree the draft level thresholds. They use these to agree a range of scores around the draft level thresholds, called the script scrutiny range. The script scrutiny range is the range of marks from which actual marked test scripts will be selected for use at the script scrutiny meeting.

Findings

We attended all three draft level setting meetings for Key Stage 2 (for mathematics, science and English). They were held in a professional manner and ample opportunity was provided for all those allowed to participate in the meeting to do so. The importance of maintaining standards was made clear at all three meetings and was the focus of discussion. Script scrutiny ranges were appropriately set. The independent adviser's clear and insightful input at these meetings enhanced the discussion and decision-making process significantly.

We noted that the reports from different TDAs were inconsistent in their presentation and analysis of data. For example, one provided a detailed analysis of the possible impact of the 2010 boycott on the end of Key Stage 2 tests that they had developed whereas another did not.

The minutes of the English draft level setting meeting were particularly detailed and provided a full and appropriate record of the meeting. This was not the case for the other draft level setting meetings where the minutes were less comprehensive, although they were made available much sooner than those for the English meeting.

Pupils operating at level 3, by their very nature, achieve a relatively small number of marks compared with their peers who attain levels 4 and 5. There are also significantly fewer of them in the in the pre-test samples since most children are operating at level 4 and above in Year 6. This relative scarcity of data and information can often make it more difficult to identify level thresholds for level 3 with the same confidence as for the other levels. This was a common issue for all three subject areas.

Recommendations / areas for improvement

We recommend that, where appropriate, greater consistency is achieved in the evidence, information and data provided in the pre-test reports and their presentation. We also recommend that omissions or areas which require clarification in the pre-test reports should be addressed before the reports are accepted and presented at meetings.

In relation to the scarcity of evidence to support level setting at level 3, we noted the decision to undertake additional analysis, where it was appropriate, to clarify outcomes and to aid decision-making. We recommend that STA considers this issue and implements any improvements to the securing of evidence for pupil performance at this level.

We recommend that all minutes provide an accurate, detailed and appropriate record of the meeting and are shared with us in a timely manner.

Script scrutiny

Script scrutiny meetings involve the marking programme leader and other senior members of the marking team for each subject (scrutineers) reviewing test scripts from the script scrutiny range. From looking at the scripts they decide where they believe the threshold mark for each level should be set. When it is not possible to decide on a definitive single mark point they will decide on a range of marks. Script scrutiny outcomes are based on the professional judgement of those present on what constitutes the mark at which each National Curriculum level covered by the test is first demonstrated.

Findings

In reviewing the record of the science script scrutiny, it appeared that the two issues identified by us in 2010 had been addressed. First, we noted that when the level threshold recommendations were shared with us, the practice of the science scrutineers (which had been to recommend a range of marks for each threshold) had been brought into line with that of mathematics and English scrutineers so that single-mark thresholds were recommended. Secondly, we were assured that for 2011, science scrutineers would have ready access to the marks awarded to the scripts on which they based their judgements, so bringing the process for science script scrutiny into line with that for English and mathematics.

Final level setting

Each subject has its own final level setting meeting at which the quantitative and qualitative data is considered. By evaluating this data, participants decide on the final thresholds for each level. As in previous years, the meetings were chaired by a senior QCDA officer and attended by QCDA staff, the TDA, the Marking Programme Leader or senior markers and the independent adviser. We, along with teacher

association representatives, attended these meeting as observers and did not participate in discussion. The thresholds agreed at the final level setting meetings were then checked and approved by the QCDA accountable officer before being released to schools.

Findings

We attended the final level setting meetings for all three subjects at Key Stage 2. They were held in a professional manner and ample opportunity was provided for all those allowed to participate in the meeting to do so. The importance of maintaining standards was made clear at all three meetings and was the focus of discussions. The percentage of pupil marks available for analysis at the time of final level setting was very high this year, allowing participants to see relatively accurately the impact of the agreed thresholds. This was an improvement on previous years and was due in large measure to the new online mark collection system.

These meetings require participants to weigh up and reconcile the available, and sometimes contradictory, quantitative and qualitative evidence and we do have questions about how consistently this issue has been addressed over time. With statistical data it is possible to calculate the level of confidence in the outcomes, but this is not the case with qualitative data and a level of confidence has to be assumed. Differences in the validity and reliability of the various sources of data and their relationship with one another make it difficult to compare them. There is a risk, explicitly recognised at one of this year's final level setting meetings, that the statistically generated threshold will be pushed to the extreme of its confidence limits to make it fit the evidence from the qualitative data, but there is no convincing argument that the qualitative data is more valid than the quantitative data nor that the statistically generated threshold lies outside the non-definable confidence limits of the qualitative data. To try to put it more simply, a case can be made for movement towards the extremes of the confidence limits of the quantitative data but not vice versa due to the absence of confidence limits for the qualitative data.

There is a decreasing number of active participants in final level setting meetings who have an overview and memory of decisions made in previous years and the rationale behind them. There is a risk that decisions may be taken which are not in line with previous good practice or which could, over time, contribute to standards shifting slightly.

As with the draft level setting meetings, we felt that the insightful input of the independent adviser enhanced discussion and improved confidence in the outcomes each of the three meetings. We wonder if the role could be enhanced to further improve confidence in the outcomes of these meetings. This could be as simple as their providing a brief written report giving an overview of the evidence and their observations and opinions.

Recommendations / areas for improvement

In order to help with the reconciliation of contradictory evidence, we recommend that STA undertakes work on the validity and reliability of different sources of evidence and considers in what circumstances one set of data might be given more weight than another.

We recommend that STA considers how the role of the independent adviser might be enhanced to improve the final level setting process.

To help ensure the maintenance of standards and to support new and existing members of staff, we recommend that STA ensures that its historical record of TDA recommended thresholds, script scrutiny recommendations and the final agreed thresholds includes a brief rationale to explain how and why decisions were made at final level setting, particularly where the evidence was not conclusive.

As new assessments are developed, STA will be considering how the standards of those assessments are best established and maintained over time. We would like to draw attention specifically to the 2011 *Regulatory Framework* which, unlike the 2010 *Code of Practice*, does not prescribe how such processes should be conducted. Hence there is now an opportunity for the relevant responsible body to think about how standard setting and maintenance could be improved in the future.

Impact

Findings: impact of changes to use of test outcomes

In 2009, full cohort testing of end of Key Stage 2 science was discontinued. Since 2010, Key Stage 2 science has been assessed at school level by teacher assessment. In order to monitor national standards, externally marked science sample tests were introduced and administered by QCDA. In 2010, individual pupil results were not reported back either to their schools or to their parents and were kept anonymous. In 2011, the arrangements were changed so that individual pupil results were sent back to schools and had to be reported to parents, but they were not made available to Ofsted or to local authorities. It is unclear what impact this change may have had on school behaviour in terms of teaching and learning and the preparation of pupils for both the pre-test and live sample test. This area needs further investigation.

Recommendations / areas for improvement

We recommend that research is undertaken to ascertain the extent to which these changes to the use of test outcomes have affected school behaviour, and indirectly, reported national standards for Key Stage 2 science in 2011.

Key Stage 2 test development: 2012 cycle

Background

This section of the report deals with the development of tests for use in 2012. It should be noted that QCDA had already made a decision to change its test development process and STA is already using this new approach for the tests to be used in 2013.

TDA's have traditionally been appointed by QCDA to develop Key Stage 2 national curriculum tests for English, mathematics and science. The tests are developed over the course of two to three years, and the process has included: writing of test items / questions, an initial formal trial of draft questions within a representative sample of schools across the country (referred to as pre-test 1), a second formal trial of the revised items / questions which are compiled to produce a set of draft test papers for each subject (referred to as pre-test 2) and finalisation of the tests and accompanying mark schemes. QCDA also appoints a modified test agency (MTA) to provide advice and develop modified versions of test papers for pupils with visual and hearing impairments. Throughout the process, there are reviews of the materials by curriculum and subject experts, teachers, markers and special educational needs (SEN) specialists. The views of pre-test administrators, teachers and pupils are considered. QCDA meets the MTA between the first and second pre-test to "establish required modifications and ensure that there is input from those with expertise in special educational needs".¹² At the end of the test development process a main test and a reserve test for each subject is produced.

Findings

In line with paragraph 57 of the *Regulatory Framework*, we kept under review key aspects of tests developed for the 2012 cycle. In particular, our review activities included a desktop review of the test developer's statistical and qualitative analysis of the pre-test 2 materials,¹³ as well as QCDA's own notes and minutes of meetings where the evidence was discussed. We also observed project boards, chaired by QCDA and attended by test developers. The purpose of the final project board was to discuss the performance of each draft test paper and to select test papers for the live test in 2012.

We also attended a meeting between the MTA and QCDA where Key Stage 2 English test papers selected for the live test in 2012 were handed over to the MTA and possible modifications for pupils with visual and hearing impairments were

¹² *Test development: level setting and maintaining standards*, QCDA, March 2010, page 11.

¹³ We were not provided with sight of mark schemes for Key Stage 2 English.

discussed. Similar meetings took place to look at the English level 6 papers and all papers for mathematics and science.

We reach our conclusions based on the above review activities and information provided by QCDA. In relation to the science tests, we note that, despite our making a request, QCDA failed to supply us with the information which we considered necessary and expedient in order to fully perform our duty to keep under review National Assessment arrangements. We also found some variability in the level of detail and quality of minutes and notes of meetings that we reviewed.

The processes and procedures for planning, designing and developing National Curriculum tests are long established. Those processes and procedures have enabled QCDA to ensure that the development of the 2012 cycle of tests run to time. We therefore commend QCDA and the TDA for their continued efforts to ensure that the tests are developed to schedule, particularly at a time of change. However we also note that there are areas for improvement. QCDA had already introduced a new test development process for the tests in 2013 and STA is currently working towards implementation of changes to National Assessment arrangements following the Government's response to the Bew Review recommendations. This period provided QCDA and is providing STA with an opportunity to undertake a detailed review of test development processes and the test papers developed.

Findings: minimising bias

Minimising bias is one of the five common criteria in the *Regulatory Framework*. It relates to the fairness of assessments for learners and is closely related to statutory equality duties. The fairness of tests depends on whether those taking the assessment can make sense of what is required of them, and are able to show what they know, understand and can do. If other factors get in the way, they can undermine the fairness of that assessment. Hence, the *Regulatory Framework* expects test developers to seek to achieve fair assessments by minimising bias.

The reports from the first and second formal trial stages demonstrate that analysis is undertaken by test developers to identify items / questions which may be potentially biased against particular groups, such as boys in comparison to girls (and vice versa). However, we are unclear about the extent of the analysis in terms of ethnic bias or any other potential bias. We note also that, in the reports, analysis around ethnicity largely focuses on the inclusion (within the pre-test sample) of pupils for whom English is an additional language and presentational issues, for example the number of images of black and minority ethnic children appearing in the test paper versus the number of images of white children, the representation of children with a range of physical disabilities and the choice of personal names used in the questions.

Recommendations / areas for improvement

We recommend that consideration is given to further ways in which National Curriculum tests could be demonstrated to be fair, for example by TDAs to carry out detailed analysis of the performance of children who fall within protected characteristics¹⁴. Whilst this may not be manageable at the pre-test stage due to the size of the sample which would be required, an analysis of outcomes after the test had been conducted could provide appropriate additional evidence.

Findings: qualitative and quantitative data

TDAs have collected qualitative as well as quantitative data about the tests in development. At various stages in the test development process, qualitative evidence is collected so that the views of external reviewers such as teachers, curriculum experts and markers are considered. During the first formal pre-test, quantitative analysis is undertaken in order to give an indication of the difficulty of individual items / questions. At the second formal trial, quantitative data is collected in order to obtain performance data about the tests as a whole in terms of the previous year's test. However, it is unclear what importance or weight is placed on the qualitative evidence, particularly where there are conflicting opinions.

Recommendations / areas for improvement

We recommend that STA provides guidance and clarification to stakeholders in relation to how quantitative and qualitative data are used to inform test development. This would aid consistency, particularly where conflicting evidence has been generated by the process, and would demonstrate how all evidence has been given due consideration when arriving at decisions.

Findings: TDA recommendations

We note that, for Key Stage 2 mathematics and English only, the TDAs included a section entitled "Recommendations" within their analysis of the second formal trial stage (pre-test 2). However, the TDAs did not set out, within their analysis (across all subjects), which of the two sets of draft test papers was more likely to achieve its specified purpose¹⁵ and meet QCDA's specification and our common criteria.¹⁶ To

¹⁴ Equalities Act 2010.

¹⁵ The specified purpose of National Curriculum tests is "ascertaining what pupils have achieved in relation to the attainment targets for that stage" (Education Act 2002, Section 76(1) (a)).

¹⁶ We set the common criteria which we use to judge the extent to which National Curriculum tests meet, or are likely to meet, their specified purpose (*Regulatory Framework for National Assessments: National Curriculum and Early Years Foundation Stage*, February 2011).

ensure transparency, it is important that the TDAs indicate, alongside their analysis, which set of draft test papers will better achieve these requirements and a rationale for their choice.

Recommendations / areas for improvement

We recognise that this particular approach will not be used in future but recommend that STA ensures that reports clearly identify which draft test papers will better achieve requirements and meet criteria.

Findings: the anchor test

Pupils sitting the draft tests during the formal trialling stages will also sit an anchor test. The anchor test is used each year to check the standard of the new test against tests in previous years. It is unclear how often the anchor test is reviewed to ensure that it reflects the current programme of study.

Recommendations / areas for improvement

Given the current National Curriculum review, we recommend that STA sets out how and when it would intend to refresh the anchor test for each subject in order to ensure that anchor tests reflect any changes to National Curriculum programmes of study.

Findings: presentation of statistical and qualitative analysis

We note minor variations (between Key Stage 2 science and Key Stage 2 English) in the level of detail contained within the test developer's analysis of the second formal trial (pre-test 2). For example, for Key Stage 2 English, helpful information is presented about data collected on the proportions of pupils who have a first language other than English, what those languages are and the pupils' level of fluency. This information is not provided for Key Stage 2 science. We accept that this variation may be due, in part, to slight differences in reporting requirements set out within the specification for each subject.

Recommendations / areas for improvement

We recommend that STA considers requiring the provision of statistical and qualitative analysis that is similar in detail and in focus of evaluation across subjects. We also recommend that STA reviews test specifications across subjects to ensure there is appropriate consistency and, where there are differences, provides a clear rationale.

Key Stage 2 English in 2012

Following the Government's acceptance of the recommendations of the Bew Review in the summer of 2011, an interim arrangement is planned for the assessment of Year 6 pupils in Key Stage 2 English in 2012. As in previous years, pupils will sit an

externally marked reading test paper. However, for writing, schools may elect that pupils sit either an externally marked writing test paper, taken during the traditional test period in May, or an internally marked writing test paper which can be used more flexibly by the school. The outcome from either format will be used by teachers as part of the evidence that will inform their summative judgement about the attainment of pupils in relation to national standards. For writing, only this teacher assessment judgement will be reported and published in summer 2012, not the test result itself, and it will “contribute towards an overall English level and ... will be used for accountability purposes” (Key Stage 2 assessment in 2010, DfE website).¹⁷ A system of external moderation in a sample of schools is also to be introduced for these statutory teacher assessment judgements of writing.

Additionally, approximately ten per cent of schools will be required to administer the externally marked test in May 2012 as part of a national sample. The outcomes of this test will be collected centrally and published “in order to monitor national standards in writing”¹⁸ but will not be used for school accountability purposes.

Findings: communicating to schools and local authorities

There is a risk that schools and local authorities may be confused by the complexity of these changes and this may impact on the management and administration of Key Stage 2 English tests in 2012. We acknowledge, however, that STA have sought to minimise this risk by putting in place strategies to communicate effectively with schools and local authorities about the changes.

Recommendations / areas for improvement

STA should continue to monitor carefully the effectiveness of communications about these arrangements and to manage this risk proactively.

Optional level 6 tests

Introduction

We have kept under review the delivery of the optional Key Stage 2 level 6 tests in English (writing and reading) and mathematics in 2011. These tests are the level 6 tests from the final cycle of the Single Level Tests pilot that were taken by pupils in the pilot schools in June 2010 before the initiative was ended later that year. They

¹⁷ www.education.gov.uk/schools/teachingandlearning/assessment/a00197755/key-stage-2-assessment-in-2012

¹⁸ www.education.gov.uk/schools/teachingandlearning/assessment/a00197755/key-stage-2-assessment-in-2012

were made available to all schools this year for optional use with their pupils and responses were internally marked.

In addition to this work we have also been keeping under review the development of the new, externally marked optional level 6 tests for delivery in May 2012 and the emerging delivery arrangements. Because the Bew Review of Key Stage 2 testing, assessment and accountability was in progress, much of this work was conducted in an uncertain environment. It was not until the Government response to the Bew Review report and recommendations on 18th July that the status of the level 6 tests in 2012 was confirmed.

Findings

QCDA has made good progress on the development and delivery arrangements for the optional level 6 tests considering the relatively late decision to make them available to schools in 2011 and the uncertainty generated both by Lord Bew's Review and the arrangements to close QCDA and establish the new executive agency of the DfE, the Standards and Testing Agency.

Because the origins of the optional level 6 tests lay in the Single Level Test pilot which was overseen by the DfE, and because it was only recently confirmed that the status of the level 6 tests will change significantly in 2012, management and governance structures have taken time to be clarified both within QCDA and between QCDA and the DfE. This has sometimes hampered communications with us and has meant that we have not always had sight of relevant information.

It is not known how many schools actually made use of the level 6 optional tests in 2011 when they were internally marked and unreported (although download statistics from the QCDA website suggest that only a minority of schools looked at the tests) but it will be important that all schools are helped to understand the curriculum upon which they draw, the expected standard in 2012 and how the tests relate to those covering levels 3-5. The context will be very different when teachers are making decisions about entering pupils for the tests in 2012 which will be externally marked (for reading and mathematics) and where outcomes will be used for school accountability purposes.

Recommendations / areas for improvement

Because many schools will not be familiar with the standards expected for pupils to succeed in the level 6 tests, nor clear about the programmes of study to which they relate, we recommend that STA provides clear guidance to help teachers and schools make appropriate decisions about which pupils are entered and understand how the tests will relate to those designed to assess at levels 3-5.

The potential impact on schools of the introduction of externally marked optional level 6 tests is unknown. In order to capture evidence of and understand this, we

recommend that STA commissions or conducts research into the impact of the level 6 tests on stakeholders.

Ongoing review activity

As outlined above, there are significant changes to statutory National Assessment arrangements at the end of Key Stage 2 in 2012 and further changes are expected in 2013 with the proposed introduction of a new test of the technical aspects of English. Our review activity will focus particularly on the implementation of these new arrangements and the implications both for the accuracy and reliability of outcomes and for comparability with results in previous years.

Key Stage 3

Background

For a number of years up to the end of 2007/8, statutory assessment at the end of Key Stage 3 consisted of tests in English (reading and writing), mathematics and science plus teacher assessment in the same subjects (including speaking and listening within English). The outcomes of both forms of assessment were reported to parents, collected by QCDA and then submitted to the DfE for publication as national statistics. For other National Curriculum subjects (known as “non-core” or “foundation” subjects), teacher assessment judgements were statutory and outcomes were also reported to parents and collected and published nationally.

From the start of the school year 2008/9, a new National Curriculum was introduced for pupils commencing Key Stage 3 in Year 7. Revised level descriptions were circulated to schools in draft form along with the new curriculum programmes of study, and these were confirmed as the versions to be used for statutory assessment by the incoming government in 2010. So, for pupils who began Key Stage 3 in September 2008 and where schools continued to offer a three-year key stage, the statutory end of Key Stage 3 assessments of summer 2011 were the first to relate to the new curriculum and revised level descriptions.

To support teachers’ judgements, QCDA was remitted to develop new resources which exemplified national standards in all foundation subjects and religious education. Materials covering levels 3 to 8 were published in 2010 and further exemplification, covering levels 1 and 2, followed early in 2011. The requirement to report teacher assessment outcomes to parents in all subjects remained in place but schools no longer had to report those for foundation subjects to the DfE via QCDA.

QCDA continued to make available optional Year 9 tests for schools to purchase if they wished along with guidance on how they could be internally administered and marked in order to generate a National Curriculum level outcome. Evidence from order numbers over the last three years suggests a decline in overall demand; a majority of schools nationally still order the tests for mathematics and science but significantly fewer order the English optional tests. We do not have evidence that allows us to say how these tests are actually used by schools and how far they may inform teacher assessment judgements.

Findings

There are no statutory moderation arrangements in place at Key Stage 3 and so it is hard for us to make any definitive statement about our confidence in the robustness and accuracy of statutory assessment outcomes. We noted an increased effort by QCDA to collect Key Stage 3 teacher assessment data. As a result of repeated contact by QCDA with schools and local authorities during the late summer of 2011,

over 99 per cent of data had been submitted by the start of the autumn term, a higher percentage than in previous years.

Ongoing review activity

As part of our duty to keep all statutory assessment arrangements under review we have undertaken a project to review current practice in teacher assessment at the end of Key Stage 3. The purpose is to gather information on how schools conduct statutory Key Stage 3 teacher assessment arrangements in a context where there are no longer statutory end of key stage tests. This is the first scoping activity that we have conducted on Key Stage 3 teacher assessment and is a step towards enabling us to review these national arrangements appropriately in the future. The project will provide an up-to-date snapshot of how schools are currently conducting statutory teacher assessment.

We have made arrangements to meet representatives from a sample of 12 schools and 8 local authorities and also from a number of stakeholder organisations to gather information and views on the practice of statutory Key Stage 3 teacher assessment. We are also asking respondents to provide their feedback on the type of approach that we might take when reviewing statutory Key Stage 3 teacher assessment. We are also conducting a literature review to explore research in this area although indications are that there is relatively little available. Alongside qualitative information, we are also bringing together quantitative information to look at patterns of teacher assessment submission and also the uptake of Year 9 optional tests.

Interim findings

Although the work has not been completed, some interim observations can be made. The practice of statutory end of Key Stage 3 teacher assessments appears to be inconsistent. Some schools have confidence in their practice, making use of support from other schools or from their local authority, whilst other schools seem less confident in their approach to teacher assessment. We have noted the varying use of pupil evidence by schools and differences in moderation activities within and across local authorities but there does appear to be a relationship between the amount of local authority / external support provided to schools and levels of confidence in teacher assessment judgements.

We have some evidence of consistent practice within schools in their approach to teacher assessment across both core and foundation subjects. We have also noted that where schools have confidence in their approach to teacher assessment, it is likely that use is made of the Assessing Pupils' Progress (APP) materials for the core subjects.

Since the end of 2008, when the requirement ceased for schools and local authorities to set targets for performance at the end Key Stage 3 and when performance tables

were no longer produced, statutory Key Stage 3 assessment appears have become a much lower priority for schools, although a few schools indicated that they were starting to use pupils' outcomes at Key Stage 3 as an indication of their likely performance in the English Baccalaureate.

We have yet to fully analyse figures from the DfE which indicate whether schools make their statutory assessment judgements at the end of Year 8 or Year 9. This should provide some indication of the extent to which pupils experience a compressed, two-year Key Stage 3 and then make an earlier start to Key Stage 4. It also seems clear, however, that many schools continue with the Key Stage 3 programme of study into Year 9 but then switch in the course of the school year; they still report Key Stage 3 assessments at the end of Year 9 even though the focus of teaching has already moved on to Key Stage 4.

The number of Year 9 optional test orders has dropped significantly in 2011 compared with last year. Some schools clearly find them useful and it still appears that a minority of secondary schools rely almost exclusively on optional tests to determine end of year and statutory end of key stage teacher assessment judgements. Others, particularly those that are practising an accelerated Key Stage 3 curriculum, seem to be paying more attention to the outcomes of the Key Stage 2 externally marked tests and use these to inform their planning towards GCSE.

As a result of recommendations in Lord Bew's Independent Review of assessment and accountability at Key Stage 2, secondary schools will need to give greater attention to the statutory, end of Key Stage 2 teacher assessment outcomes of pupils who join them at the start of Year 7. It is our objective to ensure that National Assessment arrangements give rise to outcomes which are reliable and comparable and also to promote public confidence in those National Assessment arrangements. It follows that an understanding of how teacher assessment is currently undertaken at the end of Key Stage 3 will help to inform our view of statutory teacher assessment more generally.

Once we have completed this current piece of work and considered the available evidence, we will decide what further review activity on Key Stage 3 teacher assessment would be appropriate.

Cross-key stage review activity

Parental confidence in 3 to 14 assessments

We are currently undertaking a survey to ascertain levels of parental confidence in end of stage assessments in the EYFS and in Key Stages 1, 2, and 3. The survey also aims to explore the factors that contribute to parental confidence in all four stages. Findings from the survey and any recommendations will be published on our website in early 2012 and will help us to identify future activity in this important aspect of our work.

Implications for future review activity

In 2011, almost all areas of National Assessment arrangements were subjected to review, and change, if not already in place, will impact on those arrangements in the short term. Consequently, as those changes are implemented we will keep them under review. In particular we will keep under review the forthcoming changes to the EYFS Profile, the introduction of the Year 1 Phonics Screening Check, the piloting of the Key Stage 2 test for Technical Aspects of English, changes to the overall assessment arrangements for writing at the end of Key Stage 2 and the introduction of level 6 tests for accountability.

Change in the National Assessment landscape was made more uncertain with the planned closure of QCDA, the body remitted to deliver National Assessment arrangements in 2011. We recognise the success of the thorough preparations by QCDA to deliver not only its 2011 remit but also to close down its operations and to hand them over to its successor, STA, an executive agency of the DfE. We will of course keep under review the effectiveness of the business transfer from one agency to another as STA takes on its own role and responsibilities with regard to National Assessment arrangements in 2012 and beyond.

Change also has implications for the way National Assessments are developed in order to ensure their validity. As set out in our objectives and duties, we will continue to keep under review the development of all National Assessment arrangements.

We wish to make our publications widely accessible. Please contact us if you have any specific accessibility requirements.

First published by the Office of Qualifications and Examinations Regulation in 2011

© Crown copyright 2011

You may re-use this publication (not including logos) free of charge in any format or medium, under the terms of the [Open Government Licence](#). To view this licence, [visit The National Archives](#); or write to the Information Policy Team, The National Archives, Kew, Richmond, Surrey, TW9 4DU; or email: psi@nationalarchives.gsi.gov.uk

This publication is also available on our website at www.ofqual.gov.uk

Any enquiries regarding this publication should be sent to us at:

Office of Qualifications and Examinations Regulation	
Spring Place	2nd Floor
Coventry Business Park	Glendinning House
Herald Avenue	6 Murray Street
Coventry CV5 6UB	Belfast BT1 6DN

Telephone 0300 303 3344

Textphone 0300 303 3345

Helpline 0300 303 3346