National Curriculum Assessments Review Report

2010 Key Stage 2 Tests

Contents

Executive summary

The Office of Qualifications and Examinations Regulation (Ofqual) is the regulator of qualifications, examinations and assessments in England, and is committed to ensuring that standards are maintained and that learners get the results they deserve. We keep under review the National Curriculum assessments developed by the Qualifications and Curriculum Development Agency (QCDA), in relation to a *Regulatory Framework* and *Code of Practice*.

This report presents the findings of our review of the 2010 National Curriculum key stage 2 tests, with a particular focus on the marking and standards setting processes. It considers the various meetings and other activity reviewed in 2010, and concludes as follows.

Maintenance of standards and marking

Observation of the level setting meetings, through which standards are maintained from one year to the next, and observation of the marking process, through which the standards are applied accurately and consistently to each pupil's test scripts, demonstrated that compliance with the requirements of the *2010 Code of Practice* was high. Overall, our reviewers were satisfied that standards were appropriate and were being maintained year on year, and that the marking process was robust and fit for purpose.

Introduction

Regulating National Curriculum assessments

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 statutory National Curriculum tests, and that standards are secure and consistent over time. Currently, QCDA is responsible for both the production and delivery of the tests. However, it is planned that QCDA will close in late 2011 and be replaced by a new executive agency for testing which will adopt QCDA's responsibility for the production and delivery of the tests.

For 2010, QCDA's delivery of statutory key stage 2 tests was reviewed in relation to *National Curriculum Assessments: Code of Practice 2010.* In April 2010, the Apprenticeships, Skills, Children and Learning Act (2009) established us as the independent regulator, and set two objectives for our work in keeping statutory 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.

In meeting these objectives, we are committed to ensuring that the same rigour is applied to the review of National Curriculum assessments as is brought to bear on public examinations, and that the regulation of the assessments is proportionate, accountable, consistent, transparent and targeted. We pay particular attention to those processes within the development cycle through which standards are maintained.

An overview of the level setting process

Level setting is one of the main processes through which standards in National Curriculum tests are maintained, where they are used as part of National Assessment arrangements. The level setting process involves setting threshold marks for each level of performance in the current year's tests, in order to maintain the established standard. In practical terms this means where performance of a certain standard is awarded a level in a particular year's test, that same standard of performance would be awarded the same level in any other year. The level setting process consists of three types of meeting and considers a variety of statistical and judgement-based evidence.

First, before the tests are taken by all pupils in year 6, a set of threshold marks is produced by test development agencies through statistical equating methods, using evidence from technical pre-tests of the papers to be used. These threshold marks are presented to a panel of QCDA staff at a series of draft level setting meetings. The attendees at these meetings evaluate the evidence presented by the test development agency and agree draft level thresholds. Then they agree a range of scores around the draft level thresholds, from which scripts are sampled for the script scrutiny process.

The next stage, script scrutiny meetings, is completed by comparing performance in current scripts with the established standard found at the thresholds in archive scripts. The marking programme leader and other senior members of the marking team for each test work through the script scrutiny range to decide which mark best represents the threshold standard established in previous years.

The threshold marks produced at both of these meetings are presented in turn at the final level setting meetings, which are held after the tests have been taken but before the results are returned to schools. These data allow the effect of setting the thresholds at different points to be measured against the overall performance in previous years. Additionally, the proposed thresholds are checked against the data set available for the current year to model the final data published by the Department for Education. At each of the final level setting meetings for all three subjects, a panel of senior markers, researchers, developers and QCDA staff evaluates all the evidence and comes to a decision on the final level thresholds that will be presented to QCDA's chief executive for approval.

An overview of the marking process

The marking process is designed to ensure that markers apply established national standards accurately and consistently to each pupil's script. In the autumn of the year preceding test delivery, the development of the tests is completed and the process of developing marker training material begins. This material is prepared by the marking programme leader, with support from the deputy marking programme leaders for each subject, and QCDA in close association with the test operations agency and the relevant test development agency.

Under secure conditions, the draft training material is tested on a range of markers, and revised in order to ensure that the material is robust and fit for purpose, exemplifying the mark scheme so that, after training, all markers can apply the mark scheme accurately and consistently across the range of levels.

As in previous years, the marker training programme for 2010 followed a cascade model. Typically, one marking programme leader manages four deputy marking programme leaders, who each manage and train four or five senior markers. Each senior marker is responsible for about 10 team leaders, who in turn manage and train

a team of 10 markers. Training is thus cascaded from the marking programme leader to all the markers.

All the markers have teaching experience, most at key stage 2, and are trained to ensure that they can apply the mark scheme accurately and consistently. After training, the accuracy and consistency of the marker's marking are checked at regular intervals during the marking process. This is done by setting tolerance bands to determine the acceptability of the marker's work. These bands are known as absolute mark difference (AMD) bands and are calculated by looking at the difference between the marks for the standardisation and benchmarking scripts awarded by a marker and the agreed marks awarded by the marking programme leader for the same scripts.

Using the AMD bands, markers are placed into three bands (A, B and C). Markers in band A are the most accurate and consistent. Band C markers (of whom there are very few each year) are not allowed to continue marking.

Once marking has been completed, the results and pupil scripts are returned to schools. At this point, if a school believes that a pupil or pupils have not got the result their work deserves, it may request a marking review. The marking review panel would be convened to review the marking decisions of the original marker. The panel consists of band A markers, who are retrained and standardised to ensure that they continue to apply the mark scheme accurately and consistently.

Science sampling at the end of Key Stage 2

In 2010, unlike for English and mathematics where all maintained schools were required to sit the end-of-Key-Stage-2 tests, the science end-of-Key-Stage-2 tests were only required to be taken by pupils in a national sample of schools (approximately 5 per cent). The test results were used as evidence of national standards in science. The science tests were subject to the same level setting process as for English and mathematics. Markers downloaded electronic images of the pupils' scripts to mark for science, which allowed the marker quality assurance system to be integrated with the marking of live scripts. Since the purpose of the tests was to focus on collective national standards rather than the achievement of individual pupils in the sample, no marking review facility was available to schools.

Keeping the level setting process under review

Our National Curriculum assessments team reviewed meetings during the level setting process, and considered them in relation to the *National Curriculum Assessments: Code of Practice 2010*.[1]

Reviewing the 2010 level setting process

In 2010, the National Curriculum assessments team reviewed the three types of meeting that form the level setting process. Reviewers did not participate directly in any of the meetings, but did review them for compliance with the *2010 Code of Practice*. They completed a pre-determined list of questions in order to collect evidence of compliance and to record other observations.

Sections 1, 2 and 3 of this report detail the observations made during each of the three types of meeting. Each section sets out the number of meetings that were reviewed, and then moves on to detail those instances where issues relating to compliance with the *2010 Code of Practice* were recorded. The sections conclude with areas for improvement in relation to each type of meeting. These observations cover positive findings and improved practice recorded during the reviewing, as well as issues that could, if they are not addressed, affect the future security of thresholds.

Keeping the marking process under review

Our National Curriculum assessments team reviewed meetings during the marking process and considered them in relation to the *National Curriculum Assessments: Code of Practice 2010*

Reviewing the 2010 marking process

In 2010, markers were successfully recruited and trained in line with the *National Curriculum Assessments: Code of Practice 2010*. We noted that QCDA and the test operations agency had put in place a number of new measures that ensured the smooth and efficient delivery of the training programme for markers. Key successes are noted within sections 4 and 5. In 2010, our activity focused on English. However, reviewers also sampled science and mathematics. Reviewers did not participate directly in any of the meetings, but did review them for compliance with the *2010 Code of Practice*. They completed a pre-determined list of questions in order to collect evidence of compliance and to record other observations.

---

[1] *National Curriculum Assessments: Code of Practice 2010* (Ofqual/10/4733) was published in March 2010 and is available on our website: www.ofqual.gov.uk .

Sections 4 and 5 of this report detail the observations made in relation to the marker training and reviews processes. In particular, they identify instances where issues relating to compliance with the *2010 Code of Practice* were recorded, and areas for improvement. These observations cover positive findings and improved practice recorded during reviewing, as well as issues that could, if they are not addressed, affect the future robustness and fitness for purpose of the marking process.

Section 1: Draft level setting meetings

Compliance with the *2010 Code of Practice*

These meetings were considered against section 10a (*Draft Level Setting*) of the *2010 Code of Practice*. In all three subject meetings observed, the overall level of compliance was very high, with no breaches of the *2010 Code of Practice*.

Section 2: Script scrutiny meetings

Compliance with the *2010 Code of Practice*

These meetings were considered in relation to section 10b (*Script Scrutiny*) of the *2010 Code of Practice*. In all four meetings observed, the overall level of compliance was high, with no significant breaches of the *2010 Code of Practice*. Overall, the meetings produced a series of secure recommendations for final level setting.

Management of script scrutiny meetings

Reviewers commented that all the scrutiny meetings they had observed were well chaired by the marking programme leaders, and that attendees actively participated in meetings, ensuring that robust decisions were made. Administrative support at all the meetings was very good.

Clarification of decisions

In the English reading and writing, and mathematics script scrutiny meetings, where scrutineers judged that further evidence was required to agree a cut score, additional scripts were scrutinised in an attempt to clarify decisions. This clarification was not sought at the science script scrutiny meeting.

Areas for improvement

Printing marked science scripts from onscreen marking software

This year, for the first time, science marking was conducted onscreen using the onscreen marking (OSM) software. Due to the way that marked scripts are printed from OSM software, marks awarded are not printed next to each question but grouped together on a separate mark sheet.

This proved inconvenient for science markers, as they required marks to be easily visible in order to scrutinise the scripts effectively. The marking programme leader and some of the scrutineers expressed concern about this issue, and the time required to annotate the scripts with marks. The possibility of transcription errors was also raised.

Recommendation

QCDA should:

Ensure that all scripts printed from OSM software, or similar onscreen marking solutions, have the marks readily accessible to the scrutineers. Where this involves the manual addition of marks, a quality-control procedure is required to ensure that transcription errors are not made.

Inconsistent procedures

At the script scrutiny meetings for English reading and writing, and mathematics, the scrutineers endeavoured to set a single-mark cut score in addition to a mark range. Only when this was not possible did they set just a mark range.

At the science script scrutiny meeting no attempt was made to set a single-mark cut score for any of the three levels, and only mark ranges were returned.

Recommendation

QCDA should:

Ensure that all script scrutiny meetings follow the same procedures and have similar identified outputs.

Section 3: Final level setting meetings

Compliance with the *2010 Code of Practice*

These meetings were primarily considered in relation to section 10c (*Final Level Setting*) of the *2010 Code of Practice*. All three meetings were observed, and no issues of non-compliance were found.

Final level setting meetings were firmly focused on maintaining standards

The importance of maintaining standards was evident at all three final level setting meetings. Good quality discussions and debate about the evidence presented were observed at all the meetings.

Final level setting meetings were conducted in a professional manner

Reviewers noted that final level setting meetings were held in a professional manner, and that all those able to participate in the meetings were given ample opportunity to do so.

Input from the independent adviser

Due to availability issues, a different independent adviser was present at the English and mathematics final level setting meetings than at the science sampling final level setting meeting. Both independent advisers provided clear and helpful input, which considerably enhanced the decision-making process at the meetings.

Presentation of additional data

In light of the boycott of end-of-Key-Stage-2 tests by some schools in 2010, and the fact that schools in the single level tests pilot did not participate in the key stage 2 mathematics test, additional impact data were provided at the English and mathematics final level setting meetings.

By providing impact data from 2009 that excluded these groups, where appropriate, it was possible to make realistic comparisons between the 2009 and 2010 test outcomes.

Other observations

The number of scripts available for final level setting exceeded those of previous years (English: 66.7 per cent; mathematics: 74.1 per cent; and science: 100 per cent). This increased level of availability is to be welcomed, since it addresses past concerns over the extent to which the data used at final level setting can accurately model the final data to be published by the Department for Education.

Areas for improvement

Clarification of the role of the independent adviser

Reviewers observed that, in both the English and mathematics final level setting meetings, on occasions the independent adviser was asked to suggest or agree a cut score. The role of the independent adviser is to advise the chair, whose responsibility it is to make a decision on a particular cut score.

Recommendation

QCDA should:

Clarify the role of the independent adviser at final level setting meetings.

Context of data

As noted earlier, the presentation of additional impact data proved to be very useful and allowed realistic comparisons to be made. However, it was the feeling of the reviewers that the context of the impact data was not fully articulated and explored during the English and mathematics final level setting meetings.

The vast majority of pupils in 2010 were asked to sit two series of tests at the end of key stage 2 (English and mathematics), whereas in 2009 almost all pupils took all three series of tests at the end of key stage 2 (English, mathematics and science). This different level of testing, and hence, in many cases, of preparation, might have affected attainment.

The impact data were key in arriving at decisions for level 5 writing and level 4 mathematics, and, as far as we could judge, the committee in setting the standard used the data as though they were comparing like with like, which was not strictly the case.

Additionally, for the science final level setting meeting, access to similar additional 2009 impact data that just included schools in the sample may have also been useful.

Recommendation

QCDA should:

Ensure that all impact data presented at final level setting meetings are fully explained, and any limitations are clearly articulated to those present.

Section 4: Marker training

Compliance with the *2010 Code of Practice*

These observations were considered in relation to section 3 (*Test Security at Key Stage 2)* and section 8 (*Marking at Key Stage* 2) of the *2010 Code of Practice*. No significant issues of non-compliance were found.

Early involvement of marking personnel in test development and test delivery processes

Markers were involved in the test selection and test delivery processes in 2010 at an earlier stage. By 'road testing' test papers and mark schemes at draft stage, markers were able to offer valuable insights and expertise on how straightforward a test paper was to mark accurately. This informed test paper selection and the development of the mark scheme, and, at a later stage, marker training material.

Introduction of trialling of marker training, practice scripts, marker standardisation and benchmarking scripts

This process ensured that all scripts used in the marking process had been tested on a sample of markers to ensure that each one was fit for purpose.

Introduction of practice scripts

Following marker training and before marker standardisation, markers were provided with practice scripts. This allowed markers to check their understanding of the application of the mark scheme at the agreed standard, and to receive supportive feedback before completing standardisation.

Introduction of online systems

The successful introduction of onscreen marking for key stage 2 science sampling made the overall marking process more efficient. The introduction of an online marker quality assurance system for key stage 2 English and mathematics provided markers with timely feedback and improved the efficiency of the marking process.

Consistency in delivery of key messages

For Key Stage 2 English, a number of changes to the training delivery helped to reinforce key messages, aid markers' understanding and application of the mark scheme, and ensure that training could be delivered consistently through the marker cascade and in accordance with the planned agenda. Those changes included:

■ Smaller-scale training of key stage 2 English markers (already in operation for key stage 2 mathematics and science)

■ Training markers at the outset on questions/items considered to be the most challenging to mark

■ Using a standardised trainers' script, which all trainers were required to use in their training delivery. This enabled trainers to keep to time and ensured consistent messaging across all marking teams

■ Providing commentaries for training and quality assurance scripts that explicitly demonstrated how each mark had been credited, thereby showing markers a clear link between responses and the mark scheme.

Professional and collegiate approach

Reviewers noted that, across all subjects, there was a strong collegiate approach and excellent interaction between trainers and trainees.

Organisation of training meetings

All the training meetings the reviewers observed were well organised by the test operations agency and QCDA staff, with clear systems and procedures in place.

Successful delivery of science sampling

QCDA successfully organised and managed the delivery of the marking process, including marker recruitment and training. Procedures based on those developed and utilised for single level test pilots were applied to science sampling.

Areas for improvement

Diversity of markers

Reviewers noted that there was a lack of diversity within the pool of markers selected for marker training and stakeholder feedback on test development and delivery. The *2010 Code of Practice* (paragraph 221) requires that QCDA shows that the profile of markers reflects that of the teaching community, and that the profile meets statutory [equalities] requirements. However, since QCDA did not at that time hold this information, data regarding the social characteristics of markers, for example ethnicity, age group and gender, could not be made available to us.

Recommendations

QCDA should:

Demonstrate how it is working to ensure that the marking pool reflects the teaching community and meets statutory requirements.

Carry out an equality impact assessment in line with its duties as a public body, and implement any actions.

Security

Minor security breaches were noted before and during marker training. However, the reviewers also noted that at all the training sessions they had observed the importance of security and confidentiality was reinforced by trainers, the test operations agency and QCDA. There was also notable good practice in evidence, with training rooms being checked at break times by test operations agency staff (final marker training day).

Recommendations

QCDA should:

Ensure that procedures are followed and do not compromise the security of test materials.

Continue to reinforce the need for security and confidentiality at the start of every training session, and the need to regularly check the security of rooms where test and training materials are made available.

English marker training – time allocation

In 2009[2], we '… recorded that during marker training the time spent on training to mark writing was less than that spent on training to mark reading, and was rushed.'

For 2010, reviewers noted that specific attention had been paid to the timing of training activities, that overall the training was delivered to schedule and that marking personnel appeared to be confident about their understanding of the reading and writing mark scheme for English. However, reviewers did record that there continued to be some rushing of training at meetings 6 and 8 for English in a few teams.

Recommendation

QCDA should:

Continue to keep under review the structure and content of the training cascade for English.

Errors in training material

Reviewers noted that errors were found within the training material after it had been printed. There were inconsistent approaches to the handling of those errors between mathematics and English.

---

[2] (July 2010) *Monitoring Report: 2009 Key Stage 2 Tests*, p.10, Ofqual.

Recommendation

QCDA should:

Ensure that the training material is thoroughly checked prior to the print stage and that errors are addressed to ensure consistent messaging to markers.

Training venues

Overall, the reviewers found the training venues to be very good.

However, a few training venues were noted by reviewers as not being conducive to effective training.

Recommendation

QCDA should:

Continue to ensure that all rooms selected for training venues meet QCDA's specifications prior to booking.

Scheduling

There were some scheduling issues in terms of the time allowed at user acceptance testing for gathering feedback, followed by reflection. There was also a very short window for data collection and analysis between the final user acceptance testing day and the development of marker training material. However, we commend QCDA and the test operations agency on their introduction of user acceptance training. The feedback from markers provided an invaluable source of information to marking personnel and the test operations agency, and was an important factor in the success of live marker training.

Recommendation

QCDA should:

Manage the schedule to ensure that information gathered during the user acceptance testing process can be used to inform fully the development of marker training materials.

Section 5: Marker review panels

Compliance with the *2010 Code of Practice*

These observations were considered in relation to section 12 (*Review of Marking of the 2010 Code of Practice*), and no issues of non-compliance were found.

Simplification of the reviews process

From 2010, the group review service was discontinued to simplify the services offered to schools. Instead two review services were made available: individual pupil reviews[3] and clerical reviews[4]. This simplification was intended to reduce burdens on school staff who were expected to use the mark scheme to justify their review requests.

Organisation of marker review panels

Reviewers noted that there were clear systems and procedures in place for logging scripts and recording outcomes.

The simplification of the reviews process and the direct management of administrative issues by test operations agency staff enabled marker review panel members to focus on applying the mark scheme at the agreed standard. This led to an efficient reviews process with outcomes ready for return to schools sooner than anticipated.

Quality assurance

Each marker review panel member was retrained and standardised before admittance to the panel. All scripts reviewed by panel members were quarantined by the test operations agency on a daily basis until they could be sampled by a supervisory marker before being returned to schools.

Areas for improvement

Integrity of the reviews process

It was possible for the marker review panel member to identify the original marker and supervisor during the reviews process. This may, unintentionally, influence the decision-making process.

---

[3] A full review of the pupil's entire test script. This information is extracted from QCDA's *2010 Key Stage 2 Reviews Guidance for Schools*.

[4] The correction of clerical errors identified by schools on either pupils' test scripts or the online pupil results. This information is extracted from QCDA's *2010 Key Stage 2 Reviews Guidance for Schools*.

It was also possible for marking programme leaders to review requests submitted by schools against their own original marking. Whilst we acknowledge that all other review scripts were reviewed by panel members of a higher rank and from a different marking team[5], the continuance of this arrangement is likely to compromise the integrity of the reviews process.

Recommendation

QCDA should:

Ensure that marker review panel members do not review their own marking, and that the identity of the original markers and their supervisors is not available to marker review panel members.

---

[5] All marking personnel are allocated to teams, and are trained and supported by a supervisor.

Conclusions and implications for future review activity

Compliance with the *2010 Code of Practice*

Overall compliance with the *2010 Code of Practice* during the level setting and marking processes remained very high in 2010, and so the results can be taken as secure and there can be confidence that standards were maintained.

Future review activity of the level setting and marking processes

For 2011, we will be looking to QCDA to provide evidence that all aspects of the processes are delivered to time and to quality. We hope that the areas for improvement noted in this report will assist QCDA in this respect. Our National Curriculum assessments team will continue to keep under review all the final level setting meetings in 2011, due to their high profile. Additionally, we will look closely at the marking quality assurance and review marking processes. In line with our 2011 *Regulatory Framework for National Assessments*,[6] the National Curriculum assessments team will be looking to QCDA to keep them informed of outcomes associated with draft level setting, script scrutiny and marking, so that we are fully prepared for any issues that might be raised at the final level setting meetings.

We will contact QCDA in due course to confirm the programme of work for 2011, and the National Assessment arrangements will be kept under review to ensure that processes remain robust and that standards are secure so that the public can retain confidence in them.

When we report our 2011 review activity, we expect that the new executive agency for testing will attend to our findings in order to inform its production and delivery activity in relation to the tests.

---

[6] www.ofqual.gov.uk/files/2011-regulatory-framework-for-national-assessments.pdf

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

Any enquiries regarding this document/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