

For:

➤ Teachers



**Qualifications
and Curriculum
Development
Agency**

March 2010

Test development

Level setting and maintaining standards

Contents

Test development, level setting and maintaining standards	3
Stage 1: Developing test questions.....	4
Stage 2: First pre-test	6
Stage 3: Outcomes of the first pre-test	8
Stage 4: Second pre-test	9
Stage 5: Finalisation of mark schemes and tests	11
Stage 6: Handover	14
Stage 7: Test administration.....	15
Stage 8: Level setting	16
Stage 9: Distribution of level thresholds.....	18
Stage 10: Completion of marking	19
Stage 11: Evaluation	20
Stage 12: Feedback and review.....	22

Test development, level setting and maintaining standards

This section provides detailed information on how QCDA:

- develops the national curriculum tests for key stages 1 and 2
- sets level thresholds
- maintains standards year on year.

The development of national curriculum tests can be summarised in 12 stages. The following diagram shows the timescale for developing the tests and further information about each stage.

Email assessments@qcda.org.uk if you have any questions about the contents of this section or require further information about QCDA's processes and procedures in relation to:

- test development
- level setting
- maintaining standards of the test over time.

Stage 1: Developing test questions

Who writes the tests?

Development agencies with experience of writing tests are appointed to develop the tests for each subject at each key stage. The agencies' writers include experienced subject experts and classroom practitioners.

Is there a specification for each test?

Yes. The specification ensures continuity from year to year and defines:

- the length of the test
- the coverage of the programme of study
- the characteristics of individual test questions
- the mark scheme requirements
- the balance of types of questions
- access to the test, including children with English as an additional language and children with special educational needs.

Specifications are updated annually but do not change significantly from one year to the next.

Are initial ideas trialled?

The test development agency assesses ideas and possible questions in a small number of schools, including special schools and/or schools that have pupils with special educational needs. Often one small group of children can provide a sufficient response to help the test writer improve the question. These groups include children with English as an additional language.

How are teachers and other educational experts involved in the test development process?

The tests in development are pre-tested twice throughout the two- to three-year development process. Prior to the pre-tests, the QCDA convenes test review group meetings and teacher panels to provide constructive feedback on the materials. Members of these groups are selected to provide a wide variety of educational experience and expertise so that feedback relates to the suitability and accessibility of the tests is as comprehensive as possible.

QCDA (together with the test development agencies) then considers this feedback and uses it to inform their decisions on how to adapt and improve the tests before they are finalised and printed.

The input of test review groups and teacher panels in the test development process is a key part of QCDA's quality assurance. It is crucial to ensure that the tests are suitable in terms of content and context for the pupils who will sit them.

For more information or if you are interested in participating in the development of national curriculum tests and the single level tests pilot, please visit our [get involved page](#).

Stage 2: First pre-test

What is the purpose of the first pre-test?

All test development at key stages 1 and 2 requires a minimum of two pre-tests. The first of the two pre-tests has the following purposes, it:

- determines how individual children respond to each question
- ensures that all children can understand the wording of each question and that they are not misunderstanding the question
- ensures that illustrations are appropriate and supportive
- obtains reliable data about the difficulty of individual questions.

How are the pre-tests organised?

For each key stage and subject, the test development agency and QCDA staff agree:

- the number of children to be included in the pre-test
- how the questions are to be organised into booklets and presented to children
- which sub-samples should take which booklets.

The first pre-test includes at least twice as many questions as are required for the final test.

How are schools chosen to be involved?

The criteria for choosing the sample of schools involved in a first pre-test ensure that there is a geographical spread of schools, as well as a representative selection of different types and sizes of school, including special schools and schools that have children with special educational needs. The samples are drawn by specialist teams, which are separate from the test development teams, and the test materials are administered by consultants appointed by the test development agency.

Who marks the first pre-tests?

First pre-tests are marked by teams of markers appointed and trained by the test development agency. These are usually people who have marked in this capacity for a number of years. The mark scheme is still in the early phase of development, and a large sample of children's responses is included in the mark scheme for individual questions.

How is the first pre-test analysed?

The test development agency writes a formal report to QCDA about the pre-test. This includes information on:

- the time taken by children to complete the tests
- children's scores analysed against teacher assessment data
- teachers' views
- difficulties experienced by markers
- some analysis of the answers given by children.

The quantitative analysis of each question is a key part of the test development process. Its purpose is to give a clear indication of the difficulty of individual questions. Each question is assigned a facility value that shows the proportion of children attempting the question who were successful at it. In order to inform development fully, this facility value is normally broken down into a separate value for children working at each level in the national curriculum. The level is taken either from prior attainment data or from teacher assessment, whichever is deemed the more reliable.

An additional index for each question, known as the discrimination statistic, is included. This shows the match between success on individual questions and the total score for the test. Difficult questions, if they are assessing well, should be answered successfully by the children achieving the highest overall test score. A low discrimination would tend to tell us that children are answering a question correctly more or less at random and that the question should be dropped.

Stage 3: Outcomes of the first pre-test

How are the outcomes of the first pre-test actioned?

Once the analysis of the first pre-test is complete, the agency and QCDA meet to discuss the outcomes and commence work towards the second pre-test.

Since no substantial changes are possible after the second pre-test, it is important during this stage to ensure that data is fully understood and that the agency and QCDA are confident in their work putting together test booklets for the second pre-test.

The test review group, which includes teachers and key stage experts, meets to look at the summaries of the statistical analyses and suggestions for amendments to questions in light of children's responses.

In addition, teacher pre-test panels are set up to look at specific outcomes of the pre-test.

What criteria are used to select questions for the second pre-test?

The process of selecting questions for the second pre-test is complex and takes into consideration several factors. The questions selected should:

- cover a range of skills identified in the programme of study, as defined by the focus of the question
- adequately cover the key features of the texts (for the reading test)
- provide a range of difficulty with some questions that almost all children could answer correctly at the first pre-test and some that proved demanding for the most able
- function appropriately in discriminating between the performances of the more able and the less able children
- provide a range of question and response types, including some that require extended responses and some that require short or minimal responses, for example ticking boxes or quick mental recall
- in combination, produce a distribution of performance with a mean score approximately equal to those in previous years.

Stage 4: Second pre-test

Why is a second pre-test needed?

The second pre-test differs fundamentally from the first. By this stage, the test questions have been fully composed, the test papers as a whole have been constructed using data from the first pre-test to create balanced tests of demand comparable with previous years, and the mark schemes have been thoroughly developed following the analysis of children's responses in the first pre-test.

The main purpose of the second pre-test is to obtain performance data about the tests as a whole in relation to the previous year's test. For this reason, the second pre-test is scheduled as close to the statutory test week in May as possible. A sample of approximately 1,500 children takes the new test that is scheduled for use in schools the following year alongside the current year's statutory test. In practice, the two tests are usually separated by about three or four weeks.

The second pre-test is also used to monitor performance of children with English as an additional language. This analysis is designed to feed into the test development process over the long term, not necessarily directly into the tests currently under development.

How is the second pre-test administered?

The second pre-tests for key stage 2 are administered under conditions of highest security. Only employees or consultants appointed by the test development team are involved in administering the tests in schools. A nationally representative sample of schools is selected, and test booklets are taken out to the schools for administration. The booklets resemble the intended final product extremely closely in all details, although they are not in full colour. There is usually a main test and a reserve test.

The second pre-test for key stage 1 is administered by year 2 teachers who also mark a sample of the scripts. This allows the seven-year-old children to take the pre-test in a non-threatening environment and enables the mark scheme to be trialled with classroom teachers. Teachers involved in the pre-test are asked to confirm in writing that they will maintain the confidentiality of the test materials.

What analysis is carried out?

The principal means of equating the standard of the new test to the previous year's test is to equate scores obtained by the sample of children taking the two tests. For some subjects, pupils also take an anchor test. The same anchor test is used each year, and it is possible, using statistical procedures, to equate performance on the second pre-test with that on the tests of previous years.

Who marks the second pre-test?

The second pre-test is marked against the developed mark scheme. For key stage 2, the most senior external markers are involved in the marking of this pre-test. This is done for two reasons. First, it provides these markers with invaluable experience of marking the new test in conditions that are extremely close to the live test conditions. Second, it ensures that the external marking team can provide advice to the test development agency and QCDA on the effectiveness of the mark scheme before it is finalised.

For key stage 1, the pre-test scripts are marked by experienced markers, trained by the test development agency.

As with the first pre-test marking operations, any queries and guidance offered are recorded, and any responses given by children that are not covered by the mark scheme are also noted.

Stage 5: Finalisation of mark schemes and tests

What happens after the second pre-test?

The second pre-test should be as close as possible to the final versions of the tests. This means that no substantial changes to test papers or mark schemes are expected after the second pre-test.

However, minor changes can be permitted. For example, if a question has performed poorly in the second pre-test, it can be replaced by an appropriate question taken from one of the reserve test booklets that have also gone through the second pre-test.

The mark schemes are finalised against children's responses to the second pre-test.

When are the tests modified for children with visual and hearing impairments?

QCDA appoints a modified test development agency with expertise to provide advice and develop modified versions of the tests for pupils with visual and hearing impairments. This agency meets with each of the test development agencies between the first pre-test and the second pre-test to establish the required modifications and ensure that there is input from those with expertise in special educational needs.

At this stage, the test materials are handed over to the modified test development agency. Modifiers with subject and key stage expertise mark up modifications for braille, modified large print versions of test materials and versions of the mental arithmetic test for pupils with hearing impairments. The modified test development agency also provides notes for teachers administering the modified tests, replacement mark scheme pages that may be required as the result of modifications, and training and guidance materials for use with external markers.

How are the markers for key stage 2 tests trained?

Marker training material is prepared by the lead chief marker (the most senior marker) for each subject at each key stage, in close association with the test development agency, QCDA and external marking agency leading for the subject.

Having participated in the marking of the second pre-test scripts, the lead chief marker will have a good understanding of test questions, the ways children answer those questions and the design of the mark scheme. All markers are required to mark a training pack of five scripts before they attend their first training event. This provides the trainers with evidence of any inconsistencies in marking and any misunderstanding of the mark scheme.

The lead chief marker creates a first draft of this training pack by drawing on scripts completed by children in the second pre-test. The lead chief marker works closely with the test development agency in making this selection; the five training scripts may be composites, drawn from more than five children.

How is the standard of marking key stage 2 tests monitored?

The lead chief marker and lead external marking agency at this stage also use second pre-test data and experience of marking the second pre-test scripts to consider where they will advise QCDA to place tolerance bands determining the acceptability of markers' work. These bands are known as absolute mark difference (AMD) bands and are calculated by looking at the difference between the marks awarded by a marker and the marks awarded by that marker's supervisor for the same script.

This is usually calculated by reference to at least ten 'live' scripts. Using the AMD, markers are placed into three bands (A, B and C). Markers in Band A will be the most accurate and consistent. Band C markers (there are very few each year) are not allowed to continue marking.

How is the standard of key stage 1 marking monitored?

At key stage 1, marking is carried out by year 2 teachers. The accuracy of teachers' marking, and consistency of their administration of the tests, is monitored by their local authority. Each school is subject to full audit of their marking at least once every four years. The audit consists of the observation of the administration of tasks and tests in classrooms and scrutiny of a sample of the marked outcomes of each task and test.

How is the data used to inform the level setting process?

The second pre-test involves children taking the new test alongside the statutory test. The data from both tests can be used to compare the relative demands between the two tests and establish the year-on-year continuity of standards.

In January or February of each year, QCDA meets with test development agencies to set draft level thresholds. Using a range of statistical techniques, including linear or equipercentile equating methods, children's total score on one test can be related to their total score on the new test. These methods might show, for example, that a score of 80 on the existing test equated best to a score of 81 on the new test, indicating that for these children, the new test is one mark easier than the existing test.

Is the process of setting levels for key stage 1 tests and tasks the same?

No. Teachers mark the key stage 1 tests and tasks themselves, so the level thresholds must be confirmed at this stage. In addition, the key stage 1 mark schemes have to include information about the standards and responses expected for each level awarded.

Following the review of the second pre-test data, including performance in reading comprehension, compared to the anchor test, teachers meet to carry out the Angoff process, which estimates the proportion of minimally competent learners that would correctly answer the question. There is also a scrutiny of children's responses to questions before the final levels are confirmed.

Stage 6: Handover

What happens once the tests and mark schemes have been finalised?

The test materials, which until then will have been handled directly by the test development team, are formally handed over to QCDA as electronic versions of the final checked and proof-read test paper and mark schemes.

QCDA is then solely responsible for working with the Central Office of Information (COI), the government agency responsible for printing all official documents, including statutory tests, to transfer the tests to a printer's system. The tests are proof-read and subjected to a range of quality control checks as they move through a number of proof stages before being formally signed off by QCDA as ready to print.

This stage is complete once all the test materials have been printed, collated and dispatched to schools.

Stage 7: Test administration

When are the key stage 2 tests taken?

Key stage 2 tests are taken in schools during the statutory test weeks in May. The scripts are then sent by the schools to the allocated marker who, having successfully completed the marker training and passed the initial supervisor's check, begins marking the scripts.

How are the key stage 1 tests and tasks administered?

Arrangements for the key stage 1 tests and tasks are more flexible to allow teachers to take account of the young age and particular needs of children in their classes. The role of tests and tasks is to support teachers in making their final judgement of teacher assessment at the end of key stage 1.

Teachers are required to use sufficient tasks and tests for them to be confident that their final judgement is secure. Teachers can choose a combination of tests and tasks to administer from different years, except in the case of writing, where the shorter and longer tasks used must be from the same year because they represent a pair of contrasting text types.

The tests and tasks may be administered at any point between January and the end of the school year when the final teacher assessment is made. The teacher assessment is moderated through the cycle of moderation, which is mainly carried out by local authority officers. Each school is subject to a full moderation at least once every four years, and an additional sample of random moderation is carried out each year.

What special arrangements can be made?

There is a wide range of special arrangements that can be made to make the tests accessible to children. Further details of these special arrangements can be found in *Assessment and reporting arrangements* (<http://testsandexams.qcda.gov.uk/21551.aspx>).

Stage 8: Level setting

What is level setting?

Level setting is the process that determines the minimum number of marks needed to achieve a level. Threshold marks set for each subject must be in line with the national curriculum level description. This ensures standards are maintained and each pupil's achievements are awarded the appropriate level.

What is the process used for setting levels?

To ensure the standards of performance required for the award of each level are maintained consistently from year to year, QCDA has a range of procedures (as outlined in the code of practice) both statistical and judgemental. These procedures include extensive pre-testing of all test papers so that tests can be equated year to year. They also include obtaining advice from a panel of senior markers through a script scrutiny exercise that takes place each year once the marking is well under way.

What is test equating and why is it done?

Test equating is the statistical process by which scores on one test are compared to scores on another test to understand the relative difficulty of the two tests.

The pre-testing described in Stage 4 provides data that can be analysed to equate each national curriculum test to the previous tests. This is to ensure that the standards required to achieve each level are the same each year. Test equating provides information to guide the script scrutiny process and inform the final level setting meeting.

What is script scrutiny and why is it done?

Script scrutiny is the judgemental process by which performance on one test is compared to performance on another. This is to identify the scores on the second test that represent the same level of performance as that achieved on the first. A panel of senior markers, including the marking programme leader (MPL), review scripts on the threshold marks from the previous year's test to identify features of performance at each level.

They then scrutinise scripts from the current year to identify the mark points where those same characteristics are evident. The range of scripts considered for each level is determined by the test equating. The outcomes of script scrutiny are used to inform the final level setting meeting.

What is the level setting meeting and what happens there?

After script scrutiny has been carried out, a final level setting meeting is held in June, which marks the end of the level setting process. Its purpose is to decide the mark a pupil must achieve to be awarded a particular level.

The meeting is attended by key decision makers from QCDA and representatives from the test development agencies and marking community. QCDA also appoints an independent adviser to the process and the final level setting meetings are observed by representatives from the teacher associations and unions.

The information from test equating and script scrutiny is considered, and the impact of decisions on national performance is reviewed before a decision is made. The whole process is monitored by Ofqual, the regulator, to ensure that the code of practice has been followed.

What are age-standardised scores and how are they used?

Age-standardised scores are produced by QCDA for the key stage 2 English reading and mathematics tests. They convert a pupil's actual score in the test into a standardised score that takes account of the pupil's age. This information can be used by schools to understand how each pupil is performing in relation to other pupils of different ages within the year group.

Schools do not have to use the age-standardised scores and they do not affect the pupil's test level.

Stage 9: Distribution of level thresholds

How soon are the level thresholds published?

Following the final level setting meeting for key stage 2 tests, final level threshold tables are available (<http://testsandexams.qcda.gov.uk/18954.aspx>). Level threshold tables will no longer be in the form of a printed booklet.

Key stage 1 level thresholds are published at the end of the mark schemes in the teacher's guide to the tests.

Stage 10: Completion of marking

What do markers have to do once the levels have been set?

Once they have received the confirmed thresholds, markers are able to complete their marking. This includes recording the mark totals for each child, awarding a level using tables provided for this purpose, submitting a final sample of their marking for checking purposes and dispatching all materials back to schools.

Stage 11: Evaluation

What evaluation of the tests is carried out?

Each year a number of evaluation activities are planned and undertaken.

Desk analysis of tests

An analysis of the test papers and mark schemes is commissioned by QCDA. This involves a small team of subject and assessment experts, including teachers, reviewing the test materials independently of QCDA.

The resulting report provides a critique of the tests, any questions raised, validity of the materials, clarity of questions for children, design of pages and artwork, cognitive demands required by the tests and so on. The reports arising from these influence the development of future years' tests and ensure that QCDA obtains independent views on the tests.

Error analysis

Each year, QCDA commissions agencies (sometimes the test development agencies) to carry out a detailed analysis of the responses given by children in the live tests. These are written up for teachers in the annual standards reports sent to each school.

In addition, the analysis provides an essential source of information and experience about how test questions actually worked in the live tests. It includes analysis of how children's performance changes over time and advice for teachers on implications for teaching and learning.

Are teachers' views taken into account?

Yes. QCDA always carries out a range of analyses of final test results and correspondence received from teachers and others each year and meets with all those involved in the various aspects of the test development cycle. This information then feeds into planning for the next cycle.

Do markers have the opportunity to comment on the tests?

Yes. Each year QCDA sends out a questionnaire designed to obtain the views of markers on the tests and evaluate the overall effectiveness of QCDA's work on the tests and accompanying administrative systems for external marking.

Stage 12: Feedback and review

How is feedback on the test communicated to teachers and local authorities?

Insofar as the test development cycle has an end, this comes with the publication of the standards reports, the holding of conferences for local authorities and other trainers of teachers on the outcomes of the previous year's assessments and, internal to QCDA, the finalisation of the review meetings and planning for the next cycle.



About this publication

Who is it for?

This guidance is for managers and teachers delivering foundation, higher or extended projects in schools and colleges.

What is it about

This guidance offers practical guidance on delivering projects, including principles of effective practice and case studies from schools and colleges, including working in partnership with other organisations.

What is it for?

This publication will help schools and colleges deliver successful projects.

Related publications

Guidance on managing project delivery QCA/08/3670

Introduction to the foundation project QCA/08/3671

Introduction to the higher project QCA/08/3672

Introduction to the extended project QCA/08/3673

Checklist for tutors delivering projects QCA/09/4139

Available from the QCDA orderline
www.qcda.gov.uk

For more copies:

Email orderline@qcda.gov.uk
or download from www.qcda.gov.uk/orderline

Reference QCDA/10/xxxx

Contact information:

Qualifications and Curriculum Development Agency

53-55 Butts Road, Earlsdon Park, Coventry CV1 3BH

Telephone 0300 303 3010

Textphone 0300 303 3012

Fax 0300 303 3014

info@qcda.gov.uk www.qcda.gov.uk