

## Cookies on GOV.UK

We use some essential cookies to make this website work.

We'd like to set additional cookies to understand how you use GOV.UK, remember your settings and improve government services.

We also use cookies set by other sites to help us deliver content from their services.

[Accept additional cookies](#)

[Reject additional cookies](#)

[View cookies](#)

**GOV.UK**

▼ Menu

[Home](#) ▸ [Education, training and skills](#) ▸ [School curriculum](#) ▸ [Exam regulation and administration](#)  
▸ [National assessments regulation annual report 2023](#)



Corporate report

# National assessments regulation annual report 2023

Published 25 March 2024

**Applies to England**

Contents

[Executive summary](#)

[Introduction](#)

[Section A: Priorities for 2023](#)


[Section B: Monitoring in 2023](#)

[Section C: Summary of 2023 delivery](#)

[Section D: Research](#)

[Appendix 1: Analysis of the 2023 KS2 Reading Test](#)

[Appendix 2: The KS2 markers survey in 2023](#)

 [Print this page](#)

## Executive summary

Ofqual regulates all aspects of the development and delivery of national assessments. This report details our regulatory activity and findings in relation to these assessments in 2023.

Overall in 2023 national assessment delivery was successful with significant progress being made in relation to the delivery issues that occurred in 2022 (see [Ofqual's national assessment regulation annual report](#)). The Standards and Testing Agency (STA) returned more than 99.9% of key stage 2 (KS2) results to schools on time. Ofqual's analysis of KS2 marking quality showed a high degree of reliability again in 2023; the key marking reliability metric has been consistently maintained since 2018.

The test development process carried out by STA consistently demonstrated a strong focus on validity. In addition, Ofqual is satisfied that the approach STA took to maintaining standards in 2023 provides a robust link to the standards as originally set in 2016. Legislation also requires Ofqual to promote public confidence in national assessments and with this question in mind we determined it would be appropriate to undertake a more detailed analysis of the 2023 KS2 reading test. The analyses carried out by Ofqual found no evidence that the test failed to meet its stated purpose of ascertaining what pupils have achieved in relation to the attainment targets outlined in the 2014 national curriculum, and outcome data provided by STA suggests that the test was effective in differentiating across the ability range. Ofqual's analysis highlights particular areas where there are further considerations for STA as future tests are developed. Ofqual's analysis of the test, and all the associated recommendations, can be found in appendix 1.

In 2023, Ofqual conducted a survey of KS2 markers. The purposes of the survey were to understand more about markers' backgrounds and perceptions of marking, and to enable Ofqual to capture longitudinal changes in the make up or attitudes of the marking workforce over time. Further details about the survey and its findings can be found in appendix 2 of this report.

## Introduction

### About national assessments regulation

Ofqual regulates statutory early years foundation stage profile (EYFSP) assessments and statutory national curriculum assessments. The latter, which include the Reception Baseline Assessment (RBA), the Phonics Screening Check (PSC), the Multiplication Tables Check (MTC), and key stage 1 (KS1) and key stage 2 (KS2) tests, are together referred to as 'national assessments'.

Ofqual's national assessment objectives, functions, duties and powers are set out in legislation. The objectives, as set out in our [regulatory framework](#) for national assessments, are to promote standards and confidence in national assessments, and our primary function is to keep all aspects of national assessments under review. Ofqual focuses on validity, that is, the quality of assessment. In practice, this means that we seek to ensure that the results of the assessments meet their specified statutory purposes and can therefore be trusted by those who need to use them. Ofqual's regulation also seeks to provide independent assurance about the robustness of responsible bodies' processes, and to identify risks to validity that can be addressed by responsible bodies to improve the quality of assessments over time. Ofqual is accountable to Parliament, primarily via the Education Select Committee and its annual report to Parliament, rather than to government ministers.

In 2023, Ofqual fulfilled its objectives by observing, scrutinising and reporting on key aspects of assessment validity. It takes a risk-based approach, which includes focusing on those assessments that have the 'highest stakes', such as those relied upon within school accountability measures. Ofqual has a duty to report to the Secretary of State for Education if it believes there is, or is likely to be, a significant failing in national assessment arrangements in relation to achieving one or more of the specified purposes.

The primary body responsible for national curriculum assessments is the Standards and Testing Agency (STA). STA is an executive agency within the Department for Education (DfE) and contracts suppliers to help develop, deliver or monitor national assessments. Other organisations also have responsibility for some aspects of national assessments, including local authorities, schools and other parts of the DfE, for example, teams responsible for early years assessment.

## Context for 2023

2023 was the second year of statutory national assessments at KS1 and KS2 in England following the 2-year break in primary testing that was caused by the pandemic.

In 2023, DfE confirmed that data from the KS2 tests would be used for school accountability and consequently would be published on the [compare the performance of schools and colleges in England website](#) from December 2023. Primary assessment data has not been available at school level since 2019.

Due to the cancellation of primary assessments in 2020 and 2021, 2023 was the second year that Capita were responsible for managing the delivery of the test cycle for national assessments in full. Further details on the return to KS2 testing can be found in the [national assessment regulation annual report for 2022](#).

## Section A: Priorities for 2023

Ofqual's regulatory activity for 2023 was prioritised to reflect the issues that arose during the delivery of the 2022 cycle. Consequently, Ofqual conducted enhanced monitoring of risks associated with script collection and scanning, the production of standardisation materials for moderators of teacher assessments and the performance of the national assessments helpdesk.

Ofqual sought assurance about STA's effective management of risks to delivery in several ways. Ofqual requested information and relevant documentation from STA, met regularly with senior STA staff and observed a sample of STA's internal governance, test development, marker training and standards maintenance meetings.

Another priority for Ofqual in 2023 was continuation of our analysis of the quality of marking of the KS2 tests, following the introduction of new systems and processes by the contractor for the first cycle of delivery.

These priorities informed the activity Ofqual undertook in regulating national assessments this year.

## Section B: Monitoring in 2023

### The test development process

STA is responsible for developing national curriculum assessments. The development of KS2 tests is a complex process which, in its entirety, takes approximately 3 to 4 years to complete.

A detailed explanation of the stages of the development process is published by STA in the [national curriculum test handbook](#). For each subject, the tests are constructed to meet the specification laid out in their respective [test frameworks](#).

### Ofqual's observations of the test development process

Ofqual observed a sample of meetings spanning different stages of the test development process for both KS1 and KS2 tests and across the range of subjects.

In the review meetings Ofqual attended, expert scrutiny ensured that items were, for example, an appropriate fit to the curriculum while also reflecting classroom practice. There were also clear examples where expert feedback was focused on key areas such as the wording and context of questions, to ensure they were accessible and relevant to primary school pupils.

Ofqual also observed the use of trial data and pupil responses during trialling to inform key

decisions about presentation and content of assessment instruments.

## Marking key stage 2 tests

The delivery contractor for the KS2 tests is responsible for recruiting markers to complete the marking. Contracting marking experts is a key factor in ensuring the reliability of the outcomes from these assessments. Established processes are in place for the training and quality assurance of examiners during marking.

As part of marker training quality assurance, a group of markers are invited to give feedback on the materials at an early stage of development. The feedback is used to refine the content and the approach to delivering the later rounds of training. This stage is referred to as user acceptance testing (UAT).

Following UAT, marker training is delivered through a 'cascade' model, with the most senior markers being trained first and then delivering training to more junior markers. This training process, in total, runs over several months.

For KS2 tests, items or sets of items with similar marking principles are grouped into different 'segments'. The segments are classified as either 'regular' or 'specialist', depending on the level of subject expertise required to mark them. Individual markers then mark particular segments, subject to passing the qualification process for those segments.

**In 2023 across the 3 externally-marked KS2 subjects, approximately 3,500 markers were involved in marking in the region of**

# 52 million

**segments from 3.9 million scripts.**

## Ofqual's observations of the marking process

In 2023, Ofqual monitored the quality of KS2 marker training by observing a sample of the online marker training meetings that occurred between January and May for English reading, mathematics, and grammar, punctuation and spelling. Ofqual's previous [national assessments regulation annual report](#) outlined some of the benefits of training markers online as well as some of the challenges Capita faced in its first year of managing the marking of national assessments.

The marker training meetings Ofqual observed at earlier stages in the training process in 2023, when the more senior markers were being trained, appeared to function well, as did UAT.

Although there were some initial technical issues, marker training was completed on schedule.

The purpose and objectives of the training meetings were well-defined and clear to all those participating and emphasised maintaining the security and confidentiality of the test materials. The trainers Ofqual observed exhibited their subject knowledge and their familiarity with the test, its mark scheme and the specific marking principles. Ofqual observers also noted the consistency of the training delivered by the more senior markers as they trained their own teams.

Based on these observations, it is Ofqual's judgement that the processes designed to support the quality of marker training in 2023 were sufficient to support high quality and reliable marking of KS2 tests.

## Ofqual's analysis of the quality of KS2 marking

Ofqual analysed operational marking data from 2023 using the same methodology as described in our [national assessments regulation annual report in 2022](#) to enable comparisons to be made.

During live marking, markers demonstrate they are applying the mark scheme consistently by marking responses known as 'seeds'. The seeds are introduced into markers' allocations of responses at times and intervals unknown to the marker. Approximately 1 in 40 responses marked by each marker are seeds. A comparison of the mark determined by the senior marker for the seed and the mark awarded by the marker provides an assessment of the marker's performance against the pre-agreed standard.

The data arising from the operational monitoring of the quality of marking during live marking sessions is analysed. We assume that a suitable measure of consistency of marking is based on the difference between 2 marks given for a single response: that is, an analysis of the difference between the mark set by senior markers for a seed and the actual mark awarded. This methodology is intended to give an indication of the true level of agreement across all pupil responses. It relies on the assumption that responses selected as seeds are representative of all pupil responses, in terms of their difficulty to mark.

Ofqual's analysis in 2023 found that across all 3 externally marked KS2 subjects, markers agreed with the mark set by senior markers for 99.4% of the seeds.



Figure 1: Combined exact examiner agreement levels for all 3 externally marked KS2 subjects between 2016 and 2023.

Figure 1 also illustrates that the consistency of marking of KS2 national assessments in 2023 is in line with previous years. The agreement level in 2023 is identical to the values observed in 2022, 2019 and 2018.

Based on our analysis, Ofqual can conclude that in 2023, the training and quality assurance

measures for external marking of English reading, mathematics, and grammar, punctuation and spelling were effective.

## Moderation of KS1 and KS2 writing teacher assessment

Local authorities have a statutory duty to moderate teacher assessment judgements at KS1 and KS2. STA approval to moderate English writing is only granted on successful completion of a standardisation exercise.

2023 was the second year in which the creation of moderator standardisation materials for teacher assessment of writing was outsourced by STA to the Australian Council for Education Research (ACER). One of ACER's key roles is to use examples of pupils' writing to develop suitable training materials and have these available for 3 standardisation exercises (exercises 1, 2 and 3). The local authority moderators have 2 chances to pass any one of the 3 standardisation exercises. In each case, the exercises required those taking them to grade collections of pupils' writing according to one of the 3 outcomes (working towards the expected standard, working at the expected standard, and working at greater depth within the expected standard).

In 2023, the pass rates for KS1 moderators were 92% for exercise 1, 84% for exercise 2 and 81% for exercise 3. The pass rates for moderation of KS2 writing were slightly lower, with 87% passing exercise 1, 74% passing exercise 2 and 62% passing exercise 3.

Ofqual's previous [national assessments regulatory report](#) noted that in 2022 while the overall pass rate following exercise 3 was 58%, fewer than 50% of moderators successfully passed the first 2 standardisation exercises.

## Ofqual's observation of the standardisation of teacher assessment

As a result of the issues seen last year, STA introduced more quality assurance processes that approved standardisation materials during the early stages of development by ACER. These changes ensured the materials were robust and enabled moderators to demonstrate their expertise when making their judgements which then made the standardisation process more effective in 2023.

## Live tests in 2023

In 2023, the KS2 tests were administered in the week commencing 8 May. The grammar, punctuation and spelling and mathematics tests were met with very little public commentary from stakeholders. However, some concerns were raised about the reading test by teachers, headteachers and parents. Ofqual's analysis of this test is presented in Appendix 1 of this report.



## Standards maintenance

A key requirement of national assessments is that the tests allow for accurate comparisons of performance to be made over time. Consequently, the test development process for national assessments must accommodate any differences in test difficulty that arise from building tests from a new set of questions each year. Achieving this outcome requires that standards have been set and that there is a subsequent process for ensuring they are maintained.

Following a review of the primary curriculum, new [test frameworks](#) for assessing English reading, mathematics, and grammar, punctuation and spelling were introduced, and new assessments were first administered in schools in 2016. Once the new tests had been taken by pupils (in the summer of 2016), the process of setting the standard involved groups of teachers using their professional judgement to agree a 'cut score', or threshold, that reflected the level of performance the teachers expected the pupils to reasonably achieve at the end of primary education.

In the subsequent years of administering KS2 national assessments, the performance of pupils is related back to this original standard using a statistical process known as standards maintenance. In this process, the link between the standard set in 2016 and performance in subsequent live tests is achieved through the administration of an additional test, known as an anchor test, that is taken by a representative sample of the pupils that take part in national assessment trials and live tests each year. This test remains unchanged from year to year, which allows for a statistical scaling process that links the performance in each new live test to the performance of the cohort when the original standard was set in 2016.

These processes ensure that the interpretation of the outcomes of the test remain consistent over time. The thresholds on the tests from 2017 onwards are statistically equivalent to the original standard set in 2016, and they take into account any variations in the difficulty of the tests between different years. For example, the KS2 mathematics threshold of 58 in 2022 represents the same level of attainment as the threshold of 56 in 2023.

A more detailed description of this process can be found in the [national curriculum test handbook](#) and Ofqual's [national assessments regulatory report for 2017](#).

## Ofqual's observations of the standards maintenance process

The process for maintaining test standards in 2023 was based on the same assumptions and professionally recognised techniques as in previous years.

Ofqual observed the standards maintenance meetings for both KS1 and KS2 national assessments in 2023. The procedure was carried out in line with the description set out in the [national curriculum test handbook](#).

Ofqual is satisfied that the approach STA took to maintaining standards in 2023 provides a robust link to the standards as originally set in 2016.



## Operational delivery in 2023

There were 3 noteworthy events that arose during the 2023 operational delivery cycle for national assessments.

1. In March 2023, STA announced that, owing to the additional bank holiday on Monday 8 May to celebrate the King's Coronation, the return of KS2 results would move from 4 July to 11 July. The introduction of the bank holiday necessitated that the administration of the national assessments started on the Tuesday of test week rather than the Monday. STA took the decision to move return of results back one week to ensure it could deliver results on time owing to the constraints this placed on the collection of completed scripts from schools. STA acknowledged in their communication to schools the impact this change could have on their specific arrangements, and consulted with local authorities where schools finished the summer term before or soon after return of results to ensure schools could plan for return of results.
2. In May 2023, the start of marking was delayed owing to an error in the contractors' systems such that the first round of preparatory work for marking had to be restarted. Marking therefore started on 22 May, one week later than planned.
3. On 11 July, results were returned via the Primary Assessment Gateway (PAG), the online portal managed by the contractor. On the morning of 11 July, some schools and teachers reported that they were unable to access their results at 7:30am when the PAG opened. By 10:30am 13,238 schools (approximately 79% of primary schools in England) had downloaded their results.

## Ofqual's observations of national assessment delivery in 2023

Overall in 2023, national assessment delivery was successful.

Significant progress was made in relation to the issues documented in Ofqual's [national assessments regulation annual report for 2022](#). Throughout the delivery period for 2023, Ofqual sought assurances and obtained regular updates from STA on progress against key milestones.

In relation to the issues seen in 2022:

- On return of results day, STA returned more than 99.9% of results to schools, compared with over 99.5% on return of results day in 2022. This represents more than 647,000 pupil results for each of the externally marked KS2 subjects being returned on time in 2023. There were 353 pupils with missing scripts, although 60 pupils still achieved the standard based on their performance in other papers. Between 2015 and 2019 script losses had been below 200 but in 2022 they reached nearly 2,000.
- In 2023, there were considerably fewer reports of schools and teachers experiencing problems accessing the national assessments helpline during test week. 65% of calls to the helpline were answered, compared with 38% in 2022. The average wait time this year was reduced to 22

minutes, compared with 53 minutes in 2022.

- The issue with the PAG on the morning of return of results was relatively short-lived and was associated with high system demand, with few reports of issues from teachers and schools after midday.

The system issue in May that caused the delay to the start of marking did not prevent STA from completing standards maintenance procedures normally and on schedule.

## Section C: Summary of 2023 delivery

The development of assessment materials for KS1 and KS2 in 2023 followed the published processes detailed in the subject-level [test frameworks](#) and [national curriculum test handbook](#). Marker training was delivered successfully again and the quality of marking of the KS2 assessments, as measured through the use of 'seed' responses, was high across subjects and was in line with previous years.

The delivery issues seen in 2022 in the first cycle of Capita's role as delivery partner for national assessments were not repeated in 2023, indicating that improvements to key processes have been made.

Ofqual will be seeking assurance from STA, as the responsible body, about the steps that it will take with its supplier to prevent reoccurrence of a delay to marking in 2024, as this could present a risk to timely delivery of results.

Regarding the issues with the PAG in 2023, Ofqual will be seeking assurance from STA that it is satisfied that the performance of this service has been improved in advance of the peak delivery period in 2024.

The outcomes of the national assessments at KS2 and KS1 including the multiplication tables check and the phonics screening check are published by the DfE and can be found by accessing the [explore education statistics service](#).

## Section D: Research

This year Ofqual conducted a survey of the KS2 marker workforce to understand more about its background and perceptions of marking. Ofqual has been conducting a similar study of examiners (markers) in regulated general qualifications (GQ) using surveys conducted in 2013, 2018 and [2022](#). (This survey is in addition to the survey that Capita carries out with markers every year to inform lessons learned and continuous improvement).

The survey found that more than 90% of KS2 markers indicated they were confident their marking was reliable.

This corresponds with Ofqual's analysis of marking quality, which confirmed that national assessments were marked to a high level of reliability again in 2023 (see above).

More details about the survey can be found in Appendix 2 of this report.

## Appendix 1: Analysis of the 2023 KS2 Reading Test

### Context

Legislation requires Ofqual to promote public confidence in national assessments, and with this question in mind we determined it would be appropriate to undertake a more detailed analysis of the 2023 Key Stage 2 reading test.

The analyses carried out by Ofqual, described in this appendix, found no evidence that the test failed to meet its stated purpose of ascertaining what pupils have achieved in relation to the attainment targets outlined in the 2014 national curriculum, and outcome data provided by STA suggests that the test was effective in differentiating across the ability range.

However, the analysis also highlights particular areas where there are further considerations for STA as future tests are developed.

### Introduction

Ofqual's statutory objective to promote public confidence in national assessments is in largest measure met by demonstrating that the assessments are in fact valid and reliable and can therefore be trusted for their intended uses. However, public confidence can also be (positively or negatively) affected by other questions. One of these is the perception, on the part of teachers, pupils or the wider public, of the appropriateness of tests in terms of content or level of difficulty at the point at which they are taken. The more detailed analysis of the 2023 Key Stage 2 reading test is the first such in-depth analysis undertaken since 2016. The analysis of the 2016 reading test was undertaken after some questions were posed about the appropriateness of its perceived level of difficulty for pupils. Ofqual's analysis of the 2016 reading test was published in a [2017 report](#).

It is important to note that perceptions of test difficulty on the one hand, and validity and reliability on the other, are two different issues. If a test is, or is perceived to be, more difficult or easy than the same test in previous years, this does not necessarily mean that the test results are more or less valid or reliable than those in previous years. For example, albeit in a different context, an effective maintenance of standards process in general qualifications, such as GCSEs, is used to ensure that even where there are variations year on year in the difficulty of examination questions, and therefore of grade boundaries, standards can be maintained so that the same grades

represent the same performance year on year and can be reliably compared in the same qualification taken in successive years.

In setting the level of difficulty of a test taken by a large cohort with a wide range of ability, there are always trade-offs between maximising the potential for the test to differentiate across the full range of attainment and maximising accessibility. Tests that are too easy for higher attaining pupils, for example, may not yield sufficiently reliable information on the performance of those pupils.

It is important to note, in considering varying levels of test difficulty, that STA's analysis of the outcomes of the 2023 test shows that it was effective in differentiating across the ability range.

## Results of 2023 Reading test analysis

A summary of Ofqual's analysis of the 2023 reading test is presented below. It should be noted that while we discuss the following 6 areas in separate sections, in fact there are close relationships between these areas.

### 1. Test difficulty

As outlined in the main body of this report, the threshold for a KS2 test is statistically derived by linking the test to previous live tests via an anchor test. The anchor test is an additional test or set of items that is administered to a representative sample of pupils who take the live test. This process ensures that the threshold for a KS2 test is statistically equivalent to the threshold on previous tests, back to 2016 when the standard was first set.

The table below shows the test thresholds (out of 50 marks) from 2016. The lower the threshold, the higher the difficulty of the test. The 2023 test threshold was 24, indicating that its difficulty falls between the 2016 test and the tests administered since.

Year	Threshold
2016	21
2017	26
2018	28
2019	28
2022	29
2023	24

### 2. The test experience for lower attaining pupils

The test is designed in broad terms to get more difficult as it progresses, such that, on average,

items (questions) based on the first text are easier than items based on the second text, which are in turn easier than items based on the third text. This helps to build the confidence of pupils taking the test and ensures that lower attaining pupils are not immediately met with a text which is too stretching for them. An analysis of performance by text can provide insight into the experience of lower attaining pupils, whose performance in the test is likely to be more dependent on items in the first and (to an extent) the second text. This analysis has been carried out for 2016 onwards, covering tests developed under the current test framework.

The facility index is the average mark awarded to responses to a particular item, expressed as a percentage of that item's maximum mark. For one-mark items, the facility index is equivalent to the percentage of pupils who answered the item correctly. For example, a facility of 60 would mean that 60% of the cohort answered the item correctly. For multi-mark items, the facility is the mean score across the cohort expressed as a percentage of the maximum score. For example, for a 2-mark item, a mean score of 1.4 would give a facility of 70. The higher the facility, the easier pupils found the item.

The table below shows the percentage of items across the first 2 texts for which the facility was greater than 75. In 2023, 19% of the items across texts 1 and 2 had a facility greater than 75. This is lower than in previous years. The value of 75 is used here as a proxy for items that might be considered accessible for lower attaining pupils. It is important to note that there is no specific requirement in the KS2 reading test framework relating to this value.

This provides only an approximate indicator of test experience for lower attaining pupils. Analysis using different facility index cut-off points would lead to different percentages of items flagged in the tests. It is also important to remember that this part of the discussion is focused specifically on test experience for pupils, not on the process that is used to maintain standards after the test has been taken, which we have already noted has been effective in maintaining the overall standard.

<b>Year</b>	<b>Percentage of items across texts 1 and 2 with a facility above 75</b>	<b>Text 1 mean score (%)</b>	<b>Text 2 mean score (%)</b>	<b>Text 3 mean score (%)</b>
2016	22	68	47	31
2017	43	75	70	44
2018	60	82	63	55
2019	61	81	67	55
2022	56	79	71	58
2023	19	72	62	51

(All values are given to the nearest whole number)

The table above also shows the mean pupil score, as a percentage of the total marks available, for each of the 3 texts in all KS2 Reading tests from 2016 when the current framework was introduced. For example, the value of 68 for text 1 in 2016 indicates that the average score, as a percentage of the available marks in text 1, was 68. The table shows that the mean pupil

performance in 2023 for items in text 1 and text 2 was higher than 2016, but lower than the years since.

Recommendation A: The evidence suggests that lower attaining pupils, in particular, were likely to have experienced the 2023 test as more difficult than tests in the years after 2016. STA should take into account this evidence when considering the design of future assessments.

### 3. Test completion rates and text word count

To consider this question, we analysed the omission rate for items in the 2023 reading test and compared them with previous years. The omission rate for an item refers to the percentage of pupils who did not attempt to answer it.

It should be noted that omission rate data can be difficult to interpret, and conclusions should therefore be arrived at with considerable caution. The reasons for this include the difficulty of knowing to what extent pupils 'guess' answers to selected response questions, whether they run out of time or not, and in what order they tackle questions.

With these important caveats in mind, the table below shows the omission rate for the last 3 items (designed to be broadly speaking the most difficult) in each reading test since 2016.

Year	Omission rate: last item (%)	Omission rate: second last item (%)	Omission rate: third last item (%)	Word count
2016	26*	39	18*	1,787
2017	12*	11*	23	1,937
2018	14	9*	12	1,488
2019	28	14	8*	2,168
2022	21	7*	17	1,553
2023	30	23	24	2,046

\* The asterisk denotes that the item is a selected response item

(The omission rates are given to the nearest percentage point)

The table also shows the total text word count for each test since 2016. The [reading test framework](#) stipulates a required range of 1,500 to 2,300 words. At 2,046 words, the word count for 2023 was within this range, and lower than that of the 2019 test. While the word count may impact on the amount of time pupils spend reading the texts, and therefore be related to test completion rates, there are a number of other factors that are also likely to influence completion rates. These include the complexity of the texts and the nature of interactions between items and texts (for

example, the extent to which items require pupils to draw from smaller or larger sections of the text, and the extent to which pupils are specifically directed to the relevant words, sentences or paragraphs within the text).

**Recommendation B:** For assessments of this nature, it is neither surprising, nor necessarily problematic for reliable and valid results, that not all pupils will reach the end of the test. Nevertheless, there are legitimate questions to ask about the point at which the omission rate for the last items is considered too high. STA should consider this question carefully in consultation with relevant experts in future test development.

#### 4. Text appropriateness

The appropriateness of individual texts is a particularly challenging aspect of a reading assessment to appraise, due to the lack of objective measures of text appropriateness and the lack of agreement between individuals that is often evident in discussions about judging appropriateness.

The test development process includes several points, each in advance of a trial or the live test, where external experts, including teachers and headteachers, subject and inclusion specialists, review the test material. STA also employs curriculum advisors who provide a detailed and independent perspective of the materials used in the trials and live test. In addition, pupils participating in the trials provide 'enjoyment ratings' for each of the texts to which they respond. The evidence collected by STA from these processes indicated that the materials used in the 2023 live test were viewed positively by external reviewers, including teachers and other expert review panellists and pupils who participated in the trials.

It is important to recognise the many challenges in identifying texts that may be considered by some to be inappropriate. Text appropriateness is a subjective judgement, and few individual texts are likely to attract universal approval. In addition, perceptions of the appropriateness of texts may change over time, and the test development process for one test can often span several years. This is a familiar problem in the design of all reading tests, including those used internationally. For national assessments specifically, there is also a large number of requirements and constraints that must be balanced in order to achieve optimal combination of texts, such that some trade-offs are inevitable.

**Recommendation C:** There is always likely to be some public and sector debate about the topics and themes in text selection. STA has processes in place to consider feedback on text appropriateness before the live test. It should consider how these could be further strengthened, and ensure that where appropriate, legitimate post-test observations are fed into future test development.

#### 5. Question clarity, length and complexity

A common measure used to evaluate the validity of test items is the discrimination index, which can be defined as the correlation between pupils' performance on the item and their total scores on the test. The higher the discrimination value for an item, the more effective it is in discriminating



between pupils who perform well on the test and pupils who do not. The item discriminations for the 2023 test were found to be good and in line with those of previous KS2 Reading tests.

We also considered the complexity of some items, and in particular the degree of inference required to answer them. The total number of marks in the test assessing Content Domain Reference 2d in the 2023 test ('make inferences from the text / explain and justify inferences with evidence from the text') was 23. This number falls within the permitted range stipulated in the [reading test framework](#), and is only one mark more than in 3 of the previous reading tests.

We considered the extent to which questions required pupils to draw upon ideas from across the text. Such questions are more complex and likely to increase question response time but can provide good opportunities for the assessment of pupils' ability to draw inference from reading. Cognitive domain ratings can be helpful here. The cognitive domain, which forms part of the [reading test framework](#), defines the types and degree of thinking required to answer items in the test. The domain is made up of 4 strands. One of these strands, 'accessibility of target information', relates to the accessibility of the information in the text that is needed to answer the items. STA subject experts rate each test item, for each strand, on a scale of 1 to 4. For this strand, more complex items (that is, those with higher ratings) are those for which the key information in the text is not strongly located by the item, is not prominent within the text, and not limited to one or two pieces. The mean ratings for this strand were in line with those for previous KS2 reading tests.

It is important to state that analyses such as those described here cannot provide definitive evidence in relation to any concerns about the complexity of test items. These concerns are difficult to evaluate, and while the statistics and ratings discussed above may help understand the issues, they provide only a limited picture of what are very complex interactions between pupils, items and texts.

## 6. Test development

The test development process for KS2 assessments is described in detail in the [national curriculum test handbook](#). Broadly, each reading text (along with its associated items) is trialled 3 times. First, texts and items undergo a small-scale trial. In the second trial – the Item Validation Trial (IVT) – each item is answered by approximately 300 pupils. Texts and items which have been through the IVT and have been modified in the light of evidence from the IVT are then put through a third trial. This is the Technical Pretest (TPT). For the TPT, each item is answered by around 1,000 pupils.

In the TPT, texts are carefully combined to form 'packs' of texts that are intended to meet and balance a number of constraints, such as the need to a) cover a range of text types; b) ensure the total length of the texts is appropriate; c) ensure that the difficulty across the texts is appropriate for the cohort; and d) ensure that requirements stipulated in the test specification in relation to content domain coverage, cognitive domain coverage and item types are achieved.

In the representation below of a TPT, 2 different packs are trialled (identified here as Pack 1 and Pack 2). There are multiple versions of each pack. Within a pack, the texts are identical across the versions, but not all of the items are. This allows for a greater number of items to be trialled than are needed for the live test, such that the test can be constructed using the best functioning items, while also meeting the other balances and constraints outlined above.

### Typical TPT trialling model for KS2 English reading

**Pack 1**

<b>Version 1</b>	Text A	Text B	Text C
------------------	--------	--------	--------

---

<b>Version 2</b>	Text A	Text B	Text C
------------------	--------	--------	--------

---

<b>Version 3</b>	Text A	Text B	Text C
------------------	--------	--------	--------

---

<b>Version 4</b>	Text A	Text B	Text C
------------------	--------	--------	--------

---

**Pack 2**

<b>Version 1</b>	Text D	Text E	Text F
------------------	--------	--------	--------

---

<b>Version 2</b>	Text D	Text E	Text F
------------------	--------	--------	--------

---

<b>Version 3</b>	Text D	Text E	Text F
------------------	--------	--------	--------

---

<b>Version 4</b>	Text D	Text E	Text F
------------------	--------	--------	--------

---

It had been established practice in STA when constructing KS2 reading tests for the combination of (usually 3) texts which appeared in the live test to have been previously trialled together (in the same pack) at TPT.

STA determined in August 2022 that it would be preferable to be able to construct live tests using texts from different packs, all of which had been trialled, though not necessarily in combination. In making this decision, STA consulted their technical advisors and put in place appropriate checks on the potential impacts of this change of practice. Texts for the 2023 test were, for a range of sound reasons, selected using this new approach.

There are potential benefits in greater flexibility in selecting texts from a range of pre-trialled 'packs' for optimising the quality of the final test materials. However, because this approach to text selection was novel in relation to the KS2 reading test, Ofqual sought assurances from STA about the management of attendant risks.

One such risk was of a reduced level of certainty about how texts which had previously not been trialled together would perform or be experienced by those taking the test when combined for the first time in the live test.

In the event, as this summary has shown, the 2023 test still performed well as an assessment. It is impossible to be certain about the extent to which the (for national assessments) novel approach to text selection is causally associated with the features of the 2023 test which we have analysed in sections 1-5 above.

**Recommendation D: STA should continue to focus on very careful management of**

potentially higher risks in cases where texts selected for the live test have, albeit for sound reasons, not previously been trialled in combination with each other.

## Conclusion

The main purpose of the test, as outlined in the [national curriculum test handbook](#), is to ascertain what pupils have achieved in relation to the attainment targets outlined in the 2014 National Curriculum. Based on the available statistical data, the test was effective in meeting these purposes. In addition, Ofqual's sampling of relevant STA test development meetings provided evidence that test development procedures were consistent with those outlined in the [national curriculum test handbook](#).

Analyses described in this appendix have highlighted particular areas where there are further considerations for STA as future tests are developed, in particular reading tests. As stipulated in the [regulatory framework for national assessments](#), Ofqual expects responsible bodies to keep under review, and enhance where necessary, their approach to the design, development and delivery of assessments, to assure themselves that their approach remains appropriate. This should include giving due weight to insights from public debate and feedback from stakeholders and the public.

## Regulatory follow-up

In its capacity as statutory regulator of national assessments, Ofqual has a duty to keep all aspects of national assessments under review and share any relevant conclusions and findings with the relevant responsible body. The aim of this regulatory feedback is to ensure responsible bodies are able to review and where necessary improve their processes over time, and we expect them to consider our feedback and respond appropriately.

We have shared this analysis and our recommendations with STA and expect to engage with STA further in the coming year on its consideration of our recommendations and any actions it intends to take in response. Where appropriate in the interests of maintaining public confidence in national assessments, we will report on this in a future national assessments regulation report.

## Appendix 2: The KS2 markers survey in 2023

### Introduction

Part of Ofqual's remit is to undertake research to support validity and public confidence in national assessments. In 2023, Ofqual initiated a survey of the expert markers recruited by STA's delivery contractor to mark the KS2 tests. The primary objective of the survey was to understand the make-up of the KS2 marker workforce and its perceptions of marking national assessments.

There was an additional aim of providing a benchmark from which to identify any changes that may occur in the workforce over time. Ofqual therefore created a survey that will be used in a longitudinal study. In line with the approach taken for the equivalent examiner surveys conducted by Ofqual for general qualifications, the survey will be re-administered approximately every 3 or 4 years.

## Survey methodology

The survey questions were based on the [Ofqual GCSE, AS and A level examiner survey 2022](#), which has been used to characterise that workforce over time.

The Ofqual marker survey was anonymous and optional and the results from the survey should be interpreted on this basis. The data protection procedures and the privacy statements in the survey were approved through Ofqual's internal data governance process.

Capita agreed to distribute the survey to markers. The survey was open for responses between 13 July and 8 September 2023 with one email reminder sent to markers near the end of this period.

Where survey responses are presented in the form of percentages, they exclude any missing responses. Hence, we have removed from the analysis 'prefer not to say' responses or instances where no positive indication of a choice was given. All percentages cited in the report are therefore only calculated based on respondents who gave a definite response to each question. Percentages are presented rounded to the nearest whole number or single decimal when it makes better sense to do so.

The base number of valid respondents for each question is presented next to each chart. Where appropriate, the number of responses (or the lowest number of responses) analysed is given. When markers gave free text responses, these were grouped into the most commonly reported themes and these are described in the results.

Ofqual shared preliminary high-level outcomes from the marker survey with STA prior to the publication of this report, in line with our objective to provide feedback to responsible bodies and to inform management of future national assessments delivery cycles.

## Survey results

In total, 3,542 markers were invited to take part in the survey and 1,379 markers responded, giving a response rate of 39%. The comparable GCSE, AS and A level survey had a response rate of approximately 25%.

## Profile of respondents

### Question 1

Please indicate which of the following roles you had when marking KS2 tests this year.  
(1,376 responses)



Over half the respondents to the survey (54%) were specialist markers in 2023 (those possessing a greater degree of subject expertise). The distribution of respondents amongst the specific KS2 marking roles broadly reflects the whole marker cohort.

An assistant marking programme leader (AMPL) is part of the senior marking team that is responsible for overseeing marking in their subject including the quality assurance procedures. They are also responsible for developing the training prior to live marking and training senior markers. Senior markers act as the trainers and leaders of the teams of regular or specialist markers. Regular markers mark straightforward questions. Specialist markers mark more complex questions.

### Question 2

Which KS2 test paper(s) were you involved in marking this year?  
(1,379 responses)



The proportion of survey respondents from each of the KS2 subjects broadly represents the marking workforce. For example, approximately half of the 3.9 million KS2 scripts that required marking in 2023 were the mathematics papers.

## Teaching and marking experience

### Question 3

How many times have you marked KS2 tests?  
(1,376 responses)



Over half the markers (55%) had marked KS2 tests on at least 3 occasions. Just over a quarter of the workforce (27%) were new to marking national assessments in 2023.

### Question 4

Have you ever worked as a teacher?



(1,379 responses)



Almost all of the respondents indicated they were teachers. In 2023, over half (56%) of KS2 markers indicated they currently taught in primary schools. A further 23% of markers were former primary school teachers. 21% of KS2 markers were secondary school teachers.

### Question 5

In total, how many years of teaching experience do you have?

(1,379 responses)



More than 75% of markers in 2023 had 11 or more years of teaching experience.

### Question 6

Which, if any, other exams have you marked in the last 2 years?

(1,374 responses)



76% of respondents only marked KS2 tests. Of the remaining 25%, approximately half of those (12%) also marked GCSE. The remaining 13% of markers marked a mixture of other qualifications.

Question 7 and 8 of the survey concerned markers' motivation for marking. In **question 7**, markers were asked "Why did you want to become a marker?" and respondents could tick all of the options they felt applied to them (the same set of options were used in question 8, see below).



The results to question 7 show that for more than 90% of respondents, additional income was one of their motivating factors for becoming a marker. The next most popular reasons given by markers were for professional development (61%) and to learn more about the marking process (53%).

In **question 8** markers were asked "Which one of these was your main motivation for becoming a marker?" (1,361 responses)



The most popular single reason given by respondents for becoming a marker was the additional income (66%). The next most popular motivations were professional development and to help

prepare pupils for assessments. The results in question 7 and 8 indicate that additional income and professional development are key motivating factors for the KS2 markers surveyed.

In **question 9**, markers were asked “How would you describe the level of support you are given by your school to carry out your marking duties?”

In response to this question, 21% said their employer was ‘very supportive’ and a further 29% stated that their employer was ‘neither supportive nor unsupportive’. 38% of markers selected ‘not applicable’ for this question.

## Marker perception of the training and marking process

### Question 10

Thinking about your experience of marker training for KS2 tests this year, how much do you agree or disagree with the following statements?

(The lowest number of responses analysed was 1,354)



Markers’ perceptions of the training they received were broadly positive; more than 75% of markers agreed with the statements concerning the suitability of the training and the briefings about papers and mark schemes. Similarly, more than three-quarters of markers agreed that the qualification process was fair. Over 50% of markers perceived online training to be at least as effective as face-to-face training.

Respondents to this question were given the option to add free text. Just over one-third of total respondents provided comments about training. Of those comments, more than half concerned feedback on various technical issues associated with the online training systems and processes. Some markers chose to highlight a preference for face-to-face, rather than online, training.

### Question 11

Thinking about your experience of marking KS2 tests this year, how much do you agree or disagree with the following statements?

(The lowest number of responses analysed was 1,359)



Markers’ perceptions of marking were broadly positive. The majority of respondents (over 80%) agreed that they had enough support and access to their supervisors during marking, with 70% agreeing that they received useful feedback. The majority of markers also agreed that the mark scheme they applied was sufficiently detailed (80%) and clear and unambiguous (70%).

One-third of survey respondents provided free text comments in relation to marking. More than half of the comments received were focused on interactions between markers and supervisors, which emphasises the importance markers place on this relationship (see also markers’ responses to question 13 below). Some comments included feedback suggesting the



arrangements for supervisors to respond to markers with queries could be improved.

### Question 12

Thinking about your experience of marking KS2 tests this year, how much do you agree or disagree with the following statements?

(The lowest number of responses analysed was 1,377)



93% of markers were confident in their own ability to mark the tests accurately and 83% were confident the tests in their subject were accurately marked.

One fifth of the total number of survey respondents provided free text comments in relation to the question about overall marking confidence. Two themes covered a little under half of the comments received. A small number of markers noted the impact issues with marking processes and systems had on their live marking. Some markers commented on the effect quality assurance had on their confidence, although others noted the importance and rigour of the quality assurance procedures used during live marking.

### Marker satisfaction and retention

#### Question 13

In relation to marking KS2 tests, how much do you agree or disagree with the following statements?

(The lowest number of responses analysed was 1,367)



There were many areas where markers were clearly satisfied with their overall experience of marking KS2 national assessments in 2023. 80% of markers identified they had a good working relationship with their supervisors.

70% of markers reported they found their role challenging. However, some views were less positive; 60% of markers reported they found their role stressful. Nearly 40% of markers disagreed with the statement about their workload being realistic and more than 65% of markers disagreed with the statement that their pay was satisfactory.

Overall however, the majority of markers feel that their role is meaningful (greater than 75%), take pride in their role (greater than 70%) and find it enjoyable (60%); fewer than 20% of respondents indicated they were unlikely to continue marking.

In relation to the question about marker satisfaction and retention, 40% of survey respondents added free text comments. The most common theme among markers' comments (a little under half of the total comments) related to concerns about pay. These responses are therefore consistent with markers' responses to the quantitative parts of the survey.

At the end of the survey markers were given a final opportunity to add free text comments and just under half of respondents to the survey did so. Half of these comments constituted markers' feedback and concerns about the training and marking processes and systems, management and organisation of marking and communication.

## Marker demographics

### Question 14

What is your age?  
(1,379 responses)



There was a broad age range among markers who responded to the survey, with 75% indicating they were between 31 and 60 years old. This pattern is broadly similar to the [age distribution of the teaching workforce from which the majority of KS2 markers are drawn](#).

### Question 15

What is your sex?  
(1,371 responses)



The majority of markers were female, which mirrors the [make-up of the teacher workforce in state-funded nursery and primary schools](#).

### Question 16

What is your ethnic background?  
(1,351 responses)

Ethnic origin	Percentage
White	53.4%
English, Welsh, Scottish, Northern Irish or British	37.7%
Irish	1.4%
Asian or Asian British	1.2%
Indian	1%

The table gives the 5 most common responses. Markers were most likely to indicate they were white (53%) followed by English, Welsh, Scottish, Northern Irish or British (37%). Approximately 9% of markers identified as belonging to an ethnic minority group, 1.4% indicated they were Irish, 1.2% indicated they were Asian or Asian British and 1% indicated they were Indian. Around 15% of teachers in schools [identify as belonging to an ethnic minority group](#).

### Question 17

What region of the UK do you live in?

(1,361 responses)



Markers are widely distributed among the regions, with the North West supplying the largest proportion of markers (16%). A small proportion of markers live outside England.

### Question 18

Aside from your marking work, are you retired?

(1,362 responses)



The majority of markers are working, with only 14% indicating that they are retired.

## Comparisons between KS2 markers and general qualification examiners.

Many aspects of the outcomes from the KS2 marker survey are similar to the 2022 GCSE, AS and A level examiner survey. For example, the demographic profiles show similar patterns with respect to age, sex and employment status. Significant proportions of both GQ examiners (48%) and KS2 markers (67%) disagreed with the statement about being paid sufficiently for their work. Similarly, both KS2 markers (60%) and GQ examiners (just under 60%) indicated that marking was a stressful occupation. In terms of retention, 67% of KS2 markers indicated they would continue marking in future series; for GQ examiners this was 89%. Both workforces are predominantly drawn from the teaching profession but there is relatively little overlap between the two.

## Summary of results

This is the first time that Ofqual has conducted a survey of KS2 markers, and we would like to thank all those who participated. Establishing a longitudinal study to gather insights into the make-up and sentiments of markers will enable Ofqual to undertake a broad assessment of the risks to quality and reliability of the marking of KS2 tests.

It is apparent that there is a considerable amount of teaching and marking experience amongst the respondents to the survey. Virtually all the KS2 markers are or were teachers in either primary or secondary schools, with three-quarters of respondents indicating they had been teaching for more than 11 years. The majority of the markers specialise in marking KS2 tests, with more than half of the markers having marked this assessment on at least 3 occasions. Just over a quarter of the workforce were new to KS2 marking in 2023.

Markers are broadly satisfied with their role and a significant proportion intend to continue marking for the foreseeable future, indicating that a considerable amount of the 'in built' experience will be available again. There was, however, dissatisfaction expressed among those surveyed around perceptions of workload and pay. Pay was also a common theme amongst KS2 markers' free text comments.

Importantly, the vast majority of KS2 markers who responded to the survey indicated they were satisfied with the training they received, and they were also confident in their marking of KS2 assessments in 2023, supporting Ofqual's own monitoring of these processes this year. Ofqual will continue to monitor the KS2 workforce's perceptions of marking over time.

[Back to top](#)

### Is this page useful?

Yes

No

[Report a problem with this page](#)

## Services and information

[Benefits](#)

[Births, death, marriages and care](#)

[Business and self-employed](#)

[Childcare and parenting](#)

[Citizenship and living in the UK](#)

[Crime, justice and the law](#)

[Disabled people](#)

[Driving and transport](#)

## Government activity

[Departments](#)

[News](#)

[Guidance and regulation](#)

[Research and statistics](#)

[Policy papers and consultations](#)

[Transparency](#)

[How government works](#)

[Get involved](#)

[Education and learning](#)

[Employing people](#)

[Environment and countryside](#)

[Housing and local services](#)

[Money and tax](#)

[Passports, travel and living abroad](#)

[Visas and immigration](#)

[Working, jobs and pensions](#)

---

[Help](#) [Privacy](#) [Cookies](#) [Accessibility statement](#) [Contact](#) [Terms and conditions](#)

[Rhestr o Wasanaethau Cymraeg](#) [Government Digital Service](#)

All content is available under the [Open Government Licence v3.0](#), except where otherwise stated



© [Crown copyright](#)