

National curriculum tests

Key stage 2

English reading test framework

National curriculum tests from 2016

For test developers



Standards
& Testing
Agency

© Crown copyright 2015

2016 key stage 2 English reading test framework: national curriculum tests from 2016
Electronic version product code: STA/15/7341/e ISBN: 978-1-78315-826-3

OGI You may re-use this document / publication (not including logos) free of charge in any format or medium, under the terms of the Open Government Licence v3.0. To view this licence, visit www.nationalarchives.gov.uk/doc/open-government-licence/version/3 or email: psi@nationalarchives.gsi.gov.uk.

Where we have identified any third-party copyright information, you will need to obtain permission from the copyright holders concerned.

This document is available for download on the GOV.UK website at www.gov.uk/sta.

Contents

1. Overview	4
1.1 Purposes of statutory assessment	4
2. What is a test framework?	5
3. Nature of the test	6
3.1 Population to be assessed	6
3.2 Test format	6
4. Content domain	7
5. Cognitive domain	8
5.1 Descriptions of each strand of the cognitive domain	8
6. Test specification	11
6.1 Summary	11
6.2 Breadth and emphasis	11
6.3 Format of questions and responses	13
6.4 Marking and mark schemes	14
6.5 Reporting	15
6.6 Desired psychometric properties	15
6.7 Performance descriptor	16
7. Diversity and inclusion	18
7.1 Access arrangements	19
7.2 Pupils with English as an additional language (EAL)	19
Appendix: Glossary of terminology used in the test framework	20
References	21

1. Overview

This test framework is based on the national curriculum programme of study (2014) for English, introduced for teaching in schools from September 2014 and first assessed in the summer term 2016. The framework specifies the purpose, format, content and cognitive domains of the key stage 2 English reading tests; it is not designed to be used to guide teaching and learning or to inform statutory teacher assessment.

This document has been produced to aid the test development process.

1.1 Purposes of statutory assessment

The main purpose of statutory assessment is to ascertain what pupils have achieved in relation to the attainment targets outlined in the national curriculum (2014) in English reading.

The main intended uses of the outcomes as set out in the Bew Report and the Government's consultation document on primary assessment and accountability are to:

- hold schools accountable for the attainment and progress made by their pupils
- inform parents and secondary schools about the performance of individual pupils
- enable benchmarking between schools, as well as monitoring performance locally and nationally

2. What is a test framework?

The purpose of the test framework is to provide the documentation to guide the development of the tests. The framework is written primarily for those who write test materials and to guide subsequent development and test construction. It is being made available to a wider audience for reasons of openness and transparency.

Some elements of the statutory national curriculum are not possible to assess using the current form of testing; they will need to be assessed by teachers as part of their statutory assessment of the complete national curriculum.

The framework includes those parts of the programme of study as outlined in the national curriculum (2014) that will be covered in the test (the content domain). The cognitive processes associated with the measurement of English reading are also detailed in the cognitive domain.

The test framework also includes a test specification from which valid, reliable and comparable tests will be constructed each year. This includes specifics about test format, question types, response types, marking and a clear test-level reporting strategy.

By providing all of this information in a single document, the test framework answers questions about what the test will cover, and how, in a clear and concise manner. The framework does not provide information on how teachers should teach the national curriculum.

The test development process used by the Standards and Testing Agency (STA) embeds within it the generation of validity and reliability evidence through expert review and trialling. Given the nature of the evidence collected, it is not anticipated that any additional studies will be required in order to demonstrate that the tests are fit for purpose. The test framework does not provide detail of the validity and reliability of individual tests; this will be provided in the test handbook, which will be published on the DfE's website following the administration of the test.

The test framework should be used in conjunction with the national curriculum (2014) and the annual 'Assessment and reporting arrangements' (ARA) document.

3. Nature of the test

The key stage 2 English reading test forms part of the statutory assessment arrangements for pupils at the end of key stage 2.

The test is based on the relevant sections of the national curriculum statutory programme of study (2014) for English reading at key stage 2.

The test will cover the aspects of the curriculum that lend themselves to paper-based, externally marked testing.

3.1 Population to be assessed

All eligible pupils who are registered at maintained schools, special schools, or academies (including free schools) in England and are at the end of key stage 2 will be required to take the key stage 2 English reading test, unless they have taken it in the past.

Some pupils are exempt from the tests. Further details are in the ARA, which can be found on the GOV.UK website at www.gov.uk/sta.

3.2 Test format

The key stage 2 English reading test comprises a single component. The test is administered on paper and the total testing time is 60 minutes. The testing time includes time for reading and responding to questions; there will not be separate reading time.

Table 1: Format of the test

Component	Description	Number of papers	Number of marks	Timing of paper
Paper 1: English reading test	reading booklet and separate answer booklet (a selection of texts, 1500–2300 words)	1	50	60 minutes (including reading time)
	Total	1	50	60 minutes

4. Content domain

The content domain sets out the relevant elements from the national curriculum programme of study (2014) for English at key stage 2 that are assessed in the English reading test. The tests will, over time, sample from each area of the content domain.

The key stage 2 English reading tests will focus on the comprehension elements of the national curriculum.

Table 2 shows the content domain, which sets out how elements of the curriculum will be defined for test development purposes.

Table 2: Content domain relating to questions

Content domain reference	
2a	give / explain the meaning of words in context
2b	retrieve and record information / identify key details from fiction and non-fiction
2c	summarise main ideas from more than one paragraph
2d	make inferences from the text / explain and justify inferences with evidence from the text
2e	predict what might happen from details stated and implied
2f	identify / explain how information / narrative content is related and contributes to meaning as a whole
2g	identify / explain how meaning is enhanced through choice of words and phrases
2h	make comparisons within the text

5. Cognitive domain

The cognitive domain seeks to make the thinking skills and intellectual processes required for the key stage 2 English reading test explicit. Each question will be rated against the five strands of the cognitive domain listed in the tables below to provide an indication of the cognitive demand. Information on how the questions are rated is shown in section 5.1.

The cognitive domain will be used during test development to ensure comparability of demand as well as difficulty for tests of successive years.

STA considers that the text brings another dimension to the cognitive scale and is an essential feature of test and question demand.

5.1 Descriptions of each strand of the cognitive domain

In the following tables, descriptors are provided for the top and bottom of the rating scale. Judgement is then used to categorise questions appropriately on the rest of the scale. The cognitive scale presented in this framework assumes age-appropriate texts and is a stand-alone scale for national curriculum assessments at the end of key stage 2.

Each of the strands below must be considered in the context of the national curriculum for English reading at key stage 2.

5.1.1 Accessibility of the target information

This strand relates to the accessibility of the target information that is needed to answer the question. This means:

- the number and proximity of features that need to be located in the text
- the extent to which the location of the information within the text is identified in the question
- the extent to which competing information in the text and / or distractors may mistakenly be selected

It can be thought of as, 'Where can the information be found?'

Table 3: Strand A - Accessibility of the target information

A1 (Low)	A2	A3	A4 (High)
The information that needs to be located is basic, highly prominent and limited to one or two pieces. It is clearly located by question wording and limited to a short section of the text. Competing information is limited.			The target information is not strongly located by the question. It is not prominent within the text, and not limited to one or two pieces. There is significant competing information, either within the text or in the form of functional distractors.

5.1.2 Complexity of the target information

This strand relates to the complexity of the target information that is needed to answer the question. This means:

- the lexico-grammatical density of the stimulus
- the level of concreteness / abstractness of the target information
- the level of familiarity of the information needed to answer the question

It can be thought of as, 'What is the language of the text like?'

Table 4: Strand B - Complexity of the target information

B1 (Low)	B2	B3	B4 (High)
Target information has a low level of abstractness and lexico-grammatical density, is largely familiar to pupils and is easily cued by the wording of the task.			Target information has a high level of abstractness and lexico-grammatical density and a low level of familiarity. There is a low level of semantic match between task wording and relevant information in the text.

5.1.3 Task-specific complexity

This strand relates to task-specific complexity. This means:

- the degree of cognitive complexity involved in answering the question, from retrieval through to inference and higher-level skills.

It can be thought of as, 'How much work is needed to answer the question?'

Table 5: Strand C - Task-specific complexity

C1 (Low)	C2	C3	C4 (High)
Requires only simple retrieval, with little or no inference and has concrete task requirements.			There are complex inference and abstract task requirements.

5.1.4 Response strategy

This strand relates to response strategy. This means:

- the complexity of the written response required
- the extent to which pupils need to organise / structure their response

It can be thought of as, 'How easy is it to organise and present the answer?'

Table 6: Strand D - Response strategy

D1 (Low)	D2	D3	D4 (High)
Answers will be limited to a few words and will require little organisation. The structure of response required will be clearly indicated in the question or answer booklet.			Answers are extended, and require pupils to fully structure and organise their own responses.

5.1.5 Technical knowledge required

This strand relates to the technical knowledge required. This means:

- the extent of knowledge of vocabulary required by the question
- the subject-specific technical language and knowledge required that is not given in text

It can be thought of as, 'How complex is the language of the question and / or the knowledge needed to answer it?'

Table 7: Strand E - Technical knowledge required

E1 (Low)	E2	E3	E4 (High)
No complex word meanings or subject-specific technical language is required.			Knowledge of complex word meanings and subject-specific technical language is required.

6. Test specification

This section provides details of each test component.

6.1 Summary

The test comprises one component, which will be presented to pupils as a reading booklet and an answer booklet.

Table 8: Format of the test

Component	Description	Number of papers	Number of marks	Timing of component
English reading test: booklet 1	reading booklet and separate answer booklet (a selection of texts, 1500–2300 words)	1	50	60 minutes (includes reading time)
	Total	1	50	60 minutes

6.2 Breadth and emphasis

The content and cognitive domains for the English reading tests are specified in sections 4 and 5. The test will sample from the content domain in any given year. Although every element may not be included within each test, the full range of content detailed in this document will be assessed over time. Consolidation of the key stage 1 material is assumed within the key stage 2 programme of study and therefore material from key stage 1 may appear within the key stage 2 test. The questions are placed in order of difficulty, where possible, while maintaining chronology with the text.

The following sections show the proportion of marks attributed to each of the areas of the content and cognitive domains in a test.

6.2.1 Range of texts

A range of texts will be included in the tests, including fiction, non-fiction and poetry.

Texts will be appropriate in terms of content and difficulty for pupils aged 11. This will include texts that are age-appropriate (themes in narratives will be familiar and non-fiction texts will be suitable for 10–11 year olds) and that require comprehension skills. The texts will be ordered by increasing reading demand within the reading booklet.

6.2.2 Profile of content domain

The national curriculum coverage within the test is specified in the content domain in section 4. The proportion of marks assessing each area of the content domain is shown in Table 9.

Table 9: Profile of marks by content area

National curriculum reference	Number of marks	Percentage of total mark
2a give / explain the meaning of words in context	5–10	10–20%
2b retrieve and record information / identify key details from fiction and non-fiction	8–25	16–50%
2c summarise main ideas from more than one paragraph	1–6	2–12%
2d make inferences from the text / explain and justify inferences with evidence from the text	8–25	16–50%
2e predict what might happen from details stated and implied	0–3	0–6%
2f identify / explain how information / narrative content is related and contributes to meaning as a whole	0–3	0–6%
2g identify / explain how meaning is enhanced through choice of words and phrases	0–3	0–6%
2h make comparisons within the text	0–3	0–6%

6.2.3 Profile of cognitive domain

The cognitive domain is specified in section 5. Taking into account the target information and the task itself, each question will be rated in terms of demand against each of the five strands of the cognitive domain.

For the cognitive strands A, B and C there will be questions across the range of demand 1 to 4, predominantly at 2 to 4.

For strand D, the range of questions across the cognitive domain will be in the following approximate proportions.

Table 10: Profile of ratios for strand D

Strand	Percentage of the total mark
D1	20–40%
D2 and D3	40–70%
D4	6–24%

For strand E, the majority of questions will be at E1 and E2.

6.3 Format of questions and responses

Table 11 shows how marks will be distributed across different mark tariffs.

Table 11: Profile of mark tariffs

Mark tariffs	Range of marks at mark tariff
1 mark questions	22–33 marks
2 mark questions	10–20 marks (5–10 questions)
3 mark questions	3–12 marks (1–4 questions)

The question types in the test will be distributed across the selected and constructed response questions in the proportions shown in the table. The range of question types are exemplified by, but not limited to, those listed in Table 12.

Table 12: Profile of marks by question type

Selected or constructed response	Question types	Example questions
Selected response 10–30%	Multiple choice	Where would you be most likely to see this text? Tick one of the options below.
	Ranking / ordering	Number the events below to show the order in which they happen in the story.
	Matching	Match the text to the purpose.
	Labelling	Label the text to show the title.
Short constructed response 40–60%	Find and copy	Find and copy one word that suggests...
	Short response	What does the bear eat?
Extended constructed response 20–40%	Open-ended response	Look at the paragraph beginning: <i>Once upon a time...</i> How does the writer increase the tension throughout this paragraph? Explain fully referring to the text in your answer.

6.4 Marking and mark schemes

The tests will be externally marked on screen by trained markers.

The mark schemes will give specific guidance for the marking of each question, together with general principles to ensure consistency of marking.

The mark schemes will provide the total number of marks available for each question and the criteria by which markers should award the marks. Where multiple correct answers are possible, examples of different types of correct answer will be given in the mark schemes. Where applicable, additional guidance will indicate minimally acceptable responses and unacceptable responses.

For all questions, the mark schemes will be developed during the test development process and will combine the expectations of experts with examples of pupils' responses obtained during trialling.

For multi-mark questions, if the correct answer is not reached and, therefore, full marks cannot be awarded, the mark scheme will provide details of how partial credit can be awarded.

The mark schemes will contain the following information:

- the question
- a content domain reference
- the mark allocation
- square bullets indicating the required responses or acceptable points
- round bullets exemplifying pupils' responses from the trials
- on 2 or 3 mark questions, the examples awarded higher marks will usually be placed before the examples awarded lower marks
- open 2 and 3 mark questions should also consider the pupil's ability to express their reasoning through developed answers

6.5 Reporting

The raw score on the test (the total marks achieved out of the 50 marks available) will be converted into a scaled score. Translating raw scores into scaled scores ensures performance can be reported on a consistent scale for all pupils. Scaled scores retain the same meaning from one year to the next. Therefore, a particular scaled score reflects the same standard of attainment in one year as in the previous year, having been adjusted for any differences in difficulty of the test.

Additionally, each pupil will receive an overall result indicating whether or not he or she has achieved the required standard on the test. A standard-setting exercise will be conducted on the first live test in 2016 in order to determine the scaled score needed for a pupil to be considered to have met the standard. This process will be facilitated by the performance descriptor in section 6.7, which defines the performance level required to meet the standard. In subsequent years, the standard will be maintained using appropriate statistical methods to translate raw scores on a new test into scaled scores with an additional judgemental exercise at the expected standard. The scaled score required to achieve the expected standard on the test will always remain the same.

6.6 Desired psychometric properties

While the focus of the outcome of the test will be whether a pupil has achieved the expected standard, the test must measure pupils' ability across the spectrum of attainment. As a result, the test must aim to minimise the standard error of measurement at every point on the reporting scale, particularly around the expected standard threshold.

The provision of a scaled score will aid in the interpretation of pupils' performance over time as the scaled score that represents the expected standard will be the same year-on-year. However, at the extremes of the scaled score distribution, as is standard practice, the scores will be truncated such that above or below a certain point, all pupils are awarded the same scaled score in order to minimise the effect for pupils at the ends of the distribution, where the test is not measuring optimally.

6.7 Performance descriptor

This performance descriptor describes the typical characteristics of pupils whose performance in the key stage 2 tests is at the threshold of the expected standard. Pupils who achieve the expected standard in the tests have demonstrated sufficient knowledge to be well-placed to succeed in the next phase of their education having studied the full key stage 2 programme of study in English. This performance descriptor will be used by teachers to set the standards on the new tests following their first administration in May 2016. It is not intended to be used to support teacher assessment, since it only reflects the elements of the programme of study that can be assessed in a paper-based test (see content domain in section 4).

6.7.1 Overview

Pupils working at the expected standard will be able to engage with all questions within the test. However, they will not always achieve full marks on each question, particularly if working at the threshold of the expected standard.

Questions will range from those requiring only simple retrieval in concrete tasks with little or no inference to those requiring complex inferences in abstract tasks. There will be a variety of question formats including selected response, short answer and extended answer, which require fully structured, organised and accurately communicated responses.

Question difficulty will be affected by the strands of the cognitive domain such as the accessibility and complexity of the target information within the reading text and the complexity of the technical language or other vocabulary in the question. This should be borne in mind when considering the remainder of this performance descriptor. Pupils working at the threshold of the expected standard may not give correct responses to questions where target information is not strongly located in the question or it has a high level of abstractness and lexico-grammatical density, or where there is significant competing information in the text that functions as a distractor. This will be true even when the performance descriptor determines that a skill should be within the pupil's capacity if working at the expected standard.

The following sections describe the typical characteristics of pupils in Year 6 working at the threshold of the expected standard. It is recognised that different pupils will exhibit different strengths, so this is intended as a general guide rather than a prescriptive list.

Pupils working at the expected standard are able to:

- show an understanding of the meaning of vocabulary in context (2a)
- retrieve key details and quotations from fiction and non-fiction to demonstrate understanding of character, events and information (2b)
- provide developed explanations for key information and events and for characters' actions and motivations (2b)
- accurately and selectively summarise main ideas, events, characters and information in fiction and non-fiction texts (2c)
- make developed inferences drawing on evidence from the text (2d)
- explain and justify inferences, providing evidence from the text to support reasoning (2d)
- make developed predictions that are securely rooted in the text (2e)
- identify / explain how information in non-fiction is related and contributes to meaning as a whole (2f)
- identify / explain how the sequence of events in narrative fiction contributes to meaning as a whole (2f)
- identify / explain how the choice of language enhances the meaning of texts (2g)
- make accurate and appropriate comparisons within texts (2h)

7. Diversity and inclusion

The Equality Act 2010 sets out the principles by which the national curriculum assessment and associated development activities are conducted. During the development of the tests, STA's test development division will make provision to overcome barriers to fair assessment for individuals and groups wherever possible.

National curriculum tests will also meet Ofqual's core regulatory criteria. One of the criteria refers to the need for assessment procedures to minimise bias: 'The assessment should minimise bias, differentiating only on the basis of each learner's ability to meet national curriculum requirements' (Regulatory framework for national assessment, published by Ofqual 2011).

The end of key stage 2 English reading test should:

- use appropriate means to allow all pupils to demonstrate their skills in reading
- provide a suitable challenge for all pupils and give every pupil the opportunity to achieve as high a standard as possible
- provide opportunities for all pupils to achieve, irrespective of gender, disability or special educational need, social, linguistic or cultural backgrounds
- use materials that are familiar to pupils and for which they are adequately prepared
- not be detrimental to pupils' self-esteem or confidence
- be free from stereotyping and discrimination in any form

The test development process uses the principles of universal design, as described in the 'Guidance on the principles of language accessibility in national curriculum assessments' (New language accessibility guidance, published by Ofqual 2012).

In order to improve general accessibility for all pupils, where possible, questions will be placed in order of difficulty. As with all national curriculum tests, attempts have been made to make the question rubric as accessible as possible for all pupils, including those who experience reading and processing difficulties, and those for whom English is an additional language, while maintaining an appropriate level of demand to adequately assess the content. This includes applying the principles of plain English and universal design wherever possible, conducting interviews with pupils, and taking into account feedback from expert reviewers.

For each test in development, expert opinions on specific questions are gathered, for example, at inclusion panel meetings, which are attended by experts and practitioners from across the fields of disabilities and special educational needs. This provides an opportunity for some questions to be amended or removed in response to concerns raised.

Issues likely to be encountered by pupils with specific learning difficulties have been considered in detail. Where possible, features of questions that lead to construct irrelevant variance (for example, question formats and presentational features) have been considered and questions have been presented in line with best practice for dyslexia and other specific learning difficulties.

7.1 Access arrangements

The full range of access arrangements applicable to key stage 2 assessments as set out in the ARA will be available to eligible pupils as required.

7.2 Pupils with English as an additional language (EAL)

Pupils with English as an additional language (EAL) should be registered for the national curriculum tests. If a pupil's limited ability to communicate in English means that he or she is unable to access the test, then they will be working below the standard of the English tests and should not take them, as set out in the ARA.

Appendix: Glossary of terminology used in the test framework

cognitive domain	<p>Cognitive processes refer to the thinking skills and intellectual processes that occur in response to a stimulus. The cognitive domain makes explicit the thinking skills associated with an assessment.</p> <p>The cognitive domain, as shown in this framework, also identifies other factors that may influence the difficulty of the questions.</p>
component	<p>A section of a test, presented to pupils as a test paper or test booklet is called a component. Some tests may have two or more components that each pupil needs to sit to complete the test.</p>
construct irrelevant variance	<p>Construct irrelevant variance is the variation in pupils' test scores that does not come from their knowledge of the content domain. It can result in pupils gaining fewer marks than their knowledge would suggest or lead to the award of more marks than their knowledge alone would deserve.</p> <p>The former can occur, for example, when questions in a mathematics test also unintentionally measure reading ability. The latter often occurs when unintended clues within questions allow pupils to answer correctly without having the required subject knowledge.</p>
content domain	<p>The body of subject knowledge to be assessed by the test</p>
distribution	<p>The range of possible scaled scores</p>
domain	<p>The codified definition of a body of skills and knowledge</p>
mark scheme	<p>The document explaining the creditworthy responses or the criteria that must be applied to award the mark for a question in the test</p>
national curriculum programme of study	<p>The national curriculum programme of study is the statutory description of subject knowledge, skills and understanding for a given key stage. The key stage 1 and 2 programmes of study are published online at: https://www.gov.uk/government/publications/national-curriculum-in-england-primary-curriculum</p>
performance descriptor	<p>A performance descriptor is a description of the typical characteristics of pupils working at a particular standard. For these tests, the performance descriptor will characterise the minimum performance required to be working at the appropriate standard for the end of the key stage.</p>

raw score	A raw score is the unmodified score achieved on a test, following marking. In the case of these tests it is the total marks achieved. For example, if a pupil scores 27 out of 60 possible marks, the raw score is 27. Raw scores are often then converted to other measures such as percentile ranks, standardised scores or grades.
scaled score	A score which has been translated from a raw score into a score on a fixed, defined scale is a scaled score. This allows performance to be reported on a consistent scale for all pupils, which retains the same meaning from one year to the next. Therefore, a particular scaled score reflects the same level of attainment in one year as in the previous year, having been adjusted for any differences in difficulty of the specific tests.
standard	The required level of attainment in order to be classified into a particular performance category
standard error of measurement	The standard error of measurement is a reliability estimate that allows the user to determine a confidence interval around a test score. It is a measure of the distribution of scores that would be attained by a pupil had that pupil taken the test repeatedly under the same conditions.
standard setting	The process of applying the standard to a particular test to determine the score required for a pupil to be classified within a particular performance category
test framework	A document that sets out the principles, rationale and key information about the test, and containing a test specification
test specification	A detailed description of what is to be included in a test in any single cycle of development
truncate	To shorten by removing ends

References

Independent review of key stage 2 testing, assessment and accountability (2011), Lord Bew.
www.gov.uk/government/collections/key-stage-2-ks2-testing-review

Lumley, Tom; Routitsky, Alla; Mendelovits, Juliette; and Ramalingam, Dara, 'The revised PISA reading item difficulty scheme', (2012). <http://research.acer.edu.au/pisa/5>



About this publication

Who is it for?

This document is primarily aimed at those responsible for developing the key stage 2 national curriculum test in English reading. It may also be of interest to schools with pupils in key stage 2 and other education professionals.

What does it cover?

Detailed information to ensure an appropriate test is developed, including the:

- content domain
- cognitive domain
- test specification
- the test performance descriptor

Related information

Visit the GOV.UK website at www.gov.uk/sta for all related information.

For more copies

Printed copies of this document are not available. It can be downloaded from the GOV.UK website at www.gov.uk/sta.