Sc

KEY STAGE 2

YEARS 3–6

2016

National curriculum assessments

Key stage 2 science sampling test framework (draft)

National curriculum tests from 2016

For test developers

Standards & Testing Agency

2016 Key stage 2 science sampling test framework:
national curriculum tests from 2016
Electronic version product code: STA/14/7104/e ISBN: 978-1-78315-346-6

Contents

1. Overview

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

These tests are used to monitor national standards using pupil-level matrix sampling. This approach uses a largely common set of questions which are securely administered to compare standards over a number of test cycles.

This document has been produced to aid the test development process and is therefore draft at this stage. Although any changes are expected to be minor, the document will be updated as required following evidence from trialling the tests. The document will be finalised and published on the Department for Education's (DfE's) website in advance of full sample test materials being released for schools in summer 2015.

1.1 Purposes of statutory assessment

The main purpose of sample statutory assessment is to:

- estimate national performance in relation to the attainment targets outlined in the national curriculum (2014) in science based on the performance of a sample of children.

The main use of the data is to:

- inform schools and other stakeholders about trends in children's performance in science.

1.2 Pupil-level matrix sampling

Pupil-level matrix sampling will use a number of different test instruments to assess the science national curriculum. The sampling tests will be securely administered by external assessors to a representative sample of children. The exact size of each sample will be determined by the data and reporting requirements.

Matrix sampling, requiring the use of several test instruments, will ensure that there is greater national curriculum coverage in each assessment cycle. In 2016, pupil-level matrix sampling will assess the 2014 curriculum using traditional paper-based tests.

The pupil-level matrix sampling approach also has scope for monitoring national performance against international benchmarks such as Trends in International Maths and Science Survey (TIMSS). The International Association for the Evaluation of Educational Achievement (IEA) has agreed that a sample of the questions used in the TIMSS survey can be used which will allow international assessment results to be linked with UK national sample testing.

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; it is not designed to be used to guide teaching and learning or to inform statutory teacher assessment.

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 science are also detailed in the cognitive domain.

Also included in the test framework is a test specification from which valid, reliable and comparable tests can be constructed each year. This includes specifics about test format, question types, response types and marking as well as 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 cover the content of the national curriculum in their teaching.

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 evidence collected, it is not anticipated that any additional studies will be required 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 science sampling test forms part of the statutory assessment arrangements for children at the end of key stage 2.

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

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

The tests will be securely administered in the selected schools by external administrators. Schools will not have access to the test papers after the administration. A set of questions will be made public after each assessment cycle, which schools can use for teaching purposes.

3.1 Population to be assessed

All children 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 eligible to be selected to take part in the key stage 2 science sampling test.

Some children are exempt from the tests. Further details will be made available to participating schools.

3.2 Test format

The science sampling test is comprised of three components, which will be presented to children as three separate papers. The test is administered on paper and the total testing time is 75 minutes. There are five different versions for each component and each child will be assigned one version of each component. The order of the components will vary for each child.

Table 1: Format of the test for each child

Component	Description	Number of papers	Number of marks	Approximate timing of component
Test paper b	Questions in a biology context	1(from 5 versions)	22	25 minutes
Test paper c	Questions in a chemistry context	1 (from 5 versions)	22	25 minutes
Test paper p	Questions in a physics context	1 (from 5 versions)	22	25 minutes
	Total	**3**	**66**	**75 minutes**

4. Content domain

The content domain draws out the relevant elements from the national curriculum (2014) programme of study for science at key stage 2 that are assessed in the science sampling tests It is intended that all assessable elements of the content domain will be assessed in each administration[1].

The content domain for biology, chemistry and physics also identifies strands of the programme of study that cannot be assessed or can only be partially assessed in the key stage tests due to the practical nature of those strands. These are marked with an asterisk (*) in tables 2, 3, 4 and 5 and summarised and explained in table 6.

In future years it is intended that STA will introduce performance assessments as part of the sampling test in order to assess these elements. The test framework will be expanded before the performance assessments are introduced.

To facilitate test development, each strand of the new programme of study has been assigned to one of the three core areas of science: biology, chemistry and physics. The suite of assessments for any one year in which matrix sampling occurs will cover these three core areas in approximately equal measure, even though there are slightly fewer strands for chemistry than there are for biology or physics. In the tables below, all the strands from the new programme of study for biology, chemistry and physics are given as it is possible to assess all areas to some extent with traditional paper-based tests. The 'Working scientifically' content has been described separately but will be assessed within the context of one of the three core areas of biology, chemistry or physics and will therefore always be classified under both areas.

1 In 2016, the assessments will focus on strands in the new national curriculum that are covered in years 5 and 6. By 2018, all children will have been learning science from the new curriculum for four years (all of key stage 2) and the assessments will cover the majority of strands from all years within the key stage 2 curriculum.

4.1 Content domain for biology

Table 2: Content domain for biology

Topic	Strand from programme of study		Year group
Animals including humans (B)	**B3e**	identify that animals, including humans, need the right types and amount of nutrition, and that they cannot make their own food; they get nutrition from what they eat	**3**
	B3f	identify that humans and some other animals have skeletons and muscles for support, protection and movement	**3**
	B4d	describe the simple functions of the basic parts of the digestive system in humans	**4**
	B4e	identify the different types of teeth in humans and their simple functions	**4**
	B4f	construct and interpret a variety of food chains, identifying producers, predators and prey	**4**
	B5c	describe the changes as humans develop to old age	**5**
	B6c	identify the main parts of the human circulatory system, and describe the functions of the heart, blood vessels and blood	**6**
	B6d	recognise the impact of diet, exercise, drugs and lifestyle on the way their bodies function	**6**
	B6e	describe the ways in which nutrients and water are transported within animals, including humans	**6**
Living things and their habitats (B)	**B4a**	recognise that living things can be grouped in a variety of ways	**4**
	B4b	explore and use classification keys to help group, identify and name a variety of living things in their local and wider environment*	**4**
	B4c	recognise that environments can change and that this can sometimes pose dangers to living things	**4**
	B5a	describe the differences in the life cycle of a mammal, an amphibian, an insect and a bird	**5**
	B5b	describe the life process of reproduction in some plants and animals	**5**

Topic	Strand from programme of study		Year group
Living things and their habitats (B) (Continued)	**B6a**	describe how living things are classified into broad groups according to common observable characteristics and based on similarities and differences, including micro-organisms, plants and animals	**6**
	B6b	give reasons for classifying plants and animals based on specific characteristics	**6**
Evolution and inheritance (B)	**B6f**	recognise that living things have changed over time and that fossils provide information about living things that inhabited the Earth millions of years ago	**6**
	B6g	recognise that living things produce offspring of the same kind, but normally offspring vary and are not identical to their parents	**6**
	B6h	identify how animals and plants are adapted to suit their environment in different ways and that adaptation may lead to evolution	**6**
Plants (B)	**B3a**	identify and describe the functions of different parts of flowering plants: roots, stem / trunk, leaves and flowers	**3**
	B3b	explore the requirements of plants for life and growth (air, light, water, nutrients from soil and room to grow) and how they vary from plant to plant	**3**
	B3c	investigate the way in which water is transported within plants	**3**
	B3d	explore the part that flowers play in the life cycle of flowering plants, including pollination, seed formation and seed dispersal	**3**

4.2 Content domain for chemistry

Table 3. Content domain for chemistry

Topic	Strand from programme of study		Year group
Rocks (C)	**C3a**	compare and group together different kinds of rocks on the basis of their appearance and simple physical properties*	**3**
	C3b	describe in simple terms how fossils are formed when things that have lived are trapped within rock	**3**
	C3c	recognise that soils are made from rocks and organic matter	**3**
States of matter (C)	**C4a**	compare and group together materials according to whether they are solids, liquids or gases	**4**
	C4b	observe that some materials change state when they are heated or cooled, and measure or research the temperature at which this happens in degrees Celsius (°C)*	**4**
	C4c	identify the part played by evaporation and condensation in the water cycle and associate the rate of evaporation with temperature	**4**
Properties and changes of materials (C)	**C5a**	compare and group together everyday materials on the basis of their properties, including hardness, solubility, transparency, conductivity (electrical and thermal), and response to magnets*	**5**
	C5b	know that some materials will dissolve in liquid to form a solution, and describe how to recover a substance from a solution	**5**
	C5c	use knowledge of solids, liquids and gases to decide how mixtures might be separated, including through filtering, sieving and evaporating	**5**
	C5d	give reasons, based on evidence from comparative and fair tests, for the particular uses of everyday materials including metals, wood and plastic	**5**
	C5e	demonstrate that dissolving, mixing and change of state are reversible changes	**5**
	C5f	explain that some changes result in the formation of new materials, and that this kind of change is not usually reversible, including changes associated with burning and the action of acid on bicarbonate of soda	**5**

4.3 Content domain for physics

Table 4. Content domain for physics

Topic	Strand from programme of study		Year group
Electricity (P)	**P4f**	identify common appliances that run on electricity	**4**
	P4g	construct a simple series electrical circuit*, identifying and naming its basic parts, including cells, wires, bulbs, switches and buzzers	**4**
	P4h	identify whether or not a lamp will light in a simple series circuit, based on whether or not the lamp is part of a complete loop with a battery	**4**
	P4i	recognise that a switch opens and closes a circuit and associate this with whether or not a lamp lights in a simple series circuit	**4**
	P4j	recognise some common conductors and insulators, and associate metals with being good conductors	**4**
	P6e	associate the brightness of a lamp or the volume of a buzzer with the number and voltage of cells used in the circuit	**6**
	P6f	compare and give reasons for variations in how components function, including brightness of bulbs, loudness of buzzers and on/ / off position of switches	**6**
	P6g	use recognised symbols when representing a simple circuit in a diagram	**6**
Forces (P)	**P5e**	explain that unsupported objects fall towards the Earth because of the force of gravity acting between the Earth and the falling object	**5**
	P5f	identify the effects of air resistance, water resistance and friction, that act between moving surfaces	**5**
	P5g	recognise that some mechanisms, including levers, pulleys and gears, allow a smaller force to have a greater effect	**5**
Forces and magnets (P)	**P3f**	compare how things move on different surfaces	**3**
	P3g	notice that some forces need contact between two objects, but magnetic forces can act at a distance	**3**
	P3h	observe how magnets attract or repel each other and attract some materials and not others	**3**

Topic	Strand from programme of study		Year group
Forces and magnets (P) (Continued)	**P3i**	compare and group together a variety of everyday materials on the basis of whether they are attracted to a magnet, and identify some magnetic materials	**3**
	P3j	describe magnets as having two poles	**3**
	P3k	predict whether two magnets will attract or repel each other, depending on which poles are facing	**3**
Light (P)	**P3a**	recognise that they need light in order to see things and that dark is the absence of light	**3**
	P3b	notice that light is reflected from surfaces	**3**
	P3c	recognise that light from the sun can be dangerous and that there are ways to protect their eyes	**3**
	P3d	recognise that shadows are formed when the light from a light source is blocked by an opaque	**3**
	P3e	find patterns in the way that the size of shadows change	**3**
	P6a	recognise that light appears to travel in straight lines	**6**
	P6b	use the idea that light travels in straight lines to explain that objects are seen because they give out or reflect light into the eye	**6**
	P6c	explain that we see things because light travels from light sources to our eyes or from light sources to objects and then to our eyes	**6**
	P6d	use the idea that light travels in straight lines to explain why shadows have the same shape as the objects that cast them	**6**
Sound (P)	**P4a**	identify how sounds are made, associating some of them with something vibrating	**4**
	P4b	recognise that vibrations from sounds travel through a medium to the ear	**4**
	P4c	find patterns between the pitch of a sound and features of the object that produced it	**4**
	P4d	find patterns between the volume of a sound and the strength of the vibrations that produced it	**4**
	P4e	recognise that sounds get fainter as the distance from the sound source increases	**4**

Topic	Strand from programme of study		Year group
Earth and space (P)	**P5a**	describe the movement of the Earth, and other planets, relative to the Sun in the solar system	**5**
	P5b	describe the movement of the Moon relative to the Earth	**5**
	P5c	describe the Sun, Earth and Moon as approximately spherical bodies	**5**
	P5d	use the idea of the Earth's rotation to explain day and night and the apparent movement of the sun across the sky	**5**

4.4 Content domain for 'Working scientifically'

'Working scientifically' replaces the area of 'scientific enquiry' in the national curriculum (1999) document. Between 2003 and 2012, 'scientific enquiry' had a major emphasis in the science national curriculum tests and in the science sampling tests, with 40 per cent of all questions in the tests being attributed to strands from this area of the programme of study. In the national curriculum (2014), 'Working scientifically' is considered an integral part of the other three core areas of science and will no longer be assessed as a separate domain. Questions will be integrated within the contexts of biology, chemistry and physics.

The content domain for 'Working scientifically' also includes strands of the programme of study that cannot be assessed or can only be partially assessed in the key stage tests due to their direct association with practical skills or experiences linked to school location. These are marked with an asterisk (*) in table 5 and summarised and explained in table 6.

Table 5. Content domain for 'Working scientifically'

Topic	Strand from programme of study		Key stage (lower / upper)
Planning	**WSLa**	Asking relevant questions and using different types of scientific enquiries to answer them*	**Lower**
	WSUa	Planning different types of scientific enquiries to answer questions, including recognising and controlling variables where necessary	**Upper**
Carrying out	**WSLb**	Setting up simple practical enquiries, and comparative and fair tests*	**Lower**
Measuring	**WSLc**	Making systematic and careful observations and, where appropriate, taking accurate measurements using standard units, using a range of equipment, including thermometers and data loggers*	**Lower**

Topic	Strand from programme of study		Key stage (lower / upper)
Measuring (Continued)	**WSUb**	Taking measurements, using a range of scientific equipment, with increasing accuracy and precision, taking repeat readings when appropriate*	**Upper**
Recording	**WSLd**	Gathering, recording, classifying and presenting data in a variety of ways to help in answering questions*	**Lower**
	WSLe	Recording findings using simple scientific language, drawings, labelled diagrams, keys, bar charts, and tables	**Lower**
	WSUc	Recording data and results of increasing complexity using scientific diagrams and labels, classification keys, tables, scatter graphs, bar and line graphs	**Upper**
Concluding	**WSLg**	Using results to draw simple conclusions, make predictions for new values, suggest improvements and raise further questions	**Lower**
	WSLh	Identifying differences, similarities or changes related to simple scientific ideas and processes	**Lower**
	WSLi	Using straightforward scientific evidence to answer questions, or to support their findings	**Lower**
	WSUf	Identifying scientific evidence that has been used to support or refute ideas or arguments	**Upper**
Reporting	**WSLf**	Reporting on findings from enquiries, including oral and written explanations, displays or presentations of results and conclusions*	**Lower**
	WSUe	Reporting and presenting findings from enquiries, including conclusions, causal relationships and explanations of and degree of trust in results, in oral and written forms such as displays and other presentations*	**Upper**
Further work	**WSUd**	Using test results to make predictions to set up further comparative and fair tests	**Upper**

4.5 Elements of the curriculum that cannot be assessed in this format

Table 6: Elements of the curriculum that cannot be entirely assessed in a paper-based test

National curriculum reference		Explanation of the parts that can and cannot be assessed in a paper based test
Years 3-4	**Years 5-6**	
Electricity - construct a simple series electrical circuit, identifying and naming its basic parts, including cells, wires, bulbs, switches and buzzers		This requires practical equipment to assess validly. However, the identification, naming and drawing of electrical components and circuits can be assessed in a paper-based test.
Rocks - compare and group together different kinds of rocks on the basis of their appearance and simple physical properties		Because classification is based on observed features of actual rock specimens, children should ideally physically handle rock specimens to ascertain the rocks' features. However, classification of rocks can be assessed using secondary data in a paper-based test.
States of matter - observe that some materials change state when they are heated or cooled, and measure or research the temperature at which this happens in degrees Celsius (°C)		Parts of this strand would be better assessed during practical investigations. For example, measuring or researching the temperature at which materials change state requires practical work and access to other information sources (such as text books and the internet). These skills cannot be fully assessed in a paper-based test. However, questions may be asked that draw upon a child's understanding, gained through experience, of heating and cooling materials.
Living things and their environments - explore and use classification keys to help group, identify and name a variety of living things in their local and wider environment		This requires access to, and interaction with the local environment. Questions about the child's wider environment, such as using classification keys to group, name or identify living things could be asked in a paper-based test.

National curriculum reference		Explanation of the parts that can and cannot be assessed in a paper based test
Years 3-4	**Years 5-6**	
	Properties and changes of materials - compare and group together everyday materials on the basis of their properties, including their hardness, solubility, transparency, conductivity (electricity and thermal), and response to magnets	Because classification is based on observed or tested properties, children need to physically handle materials to do this. Children can classifiy materials based on their properties using secondary data and can be assessed on their understanding of properties of materials in a paper-based test.
Planning - asking relevant questions and using different types of scientific enquiries to answer them		This requires a discursive approach between the child and their teacher. Children can be assessed on their understanding of what types of scientific enquiry would be appropriate for given situations or to identify a suitable question for an investigation, for a given context, in a paper-based test.
Carrying out - setting up simple practical enquiries, and comparative and fair tests		This requires the child to use practical equipment, including making measurements and being able to either follow instructions or make decisions about a procedure. This is not possible in a paper-based test. Children can identify what equipment could be used for an enquiry in a paper-based test.
Measuring - making systematic and careful observations and, where appropriate, taking accurate measurements using standard units, using a range of equipment, including thermometers and data loggers		This requires children to use actual equipment over a period of time. It also requires children to determine when readings should be taken (ie to manage their time during a practical activity). Children can be assessed on reading measurements from diagrams or photographs of equipment and also using correct units in a paper-based test.

National curriculum reference		Explanation of the parts that can and cannot be assessed in a paper based test
Years 3-4	**Years 5-6**	
	Measuring - taking measurements, using a range of scientific equipment, with increasing accuracy and precision, taking repeat readings when appropriate	This requires the child to use equipment and to determine how and when to take readings. Children can be assessed on their understanding of when and why repeat readings need to be taken in a paper-based test.
Recording - gathering, recording, classifying and presenting data in a variety of ways to help in answering questions		This requires the child to gather, process and present their own data based on a specific investigative question which is not possible in a paper-based test. Children can be assessed on their ability to use a variety of formats to record specific data provided to them in a test question and use data to answer questions relating to a given investigation in a paper-based test.
Reporting - reporting on findings from enquiries, including oral and written explanations, displays or presentations of results and conclusions		This requires the child to have carried out their own investigative work and to present information in ways that are not possible in a paper-based test.
	Reporting - reporting and presenting findings from enquiries, including conclusions, causal relationships and explanations of and degree of trust in results, in oral and written forms such as displays and other presentations	This strand would take too long and some activities (eg giving a presentation) would not be possible in a test situation. Children can be asked to write conclusions, describe and / or explain causal relationships and to provide scientific explanations for a given set of data in a paper-based test.

5. Cognitive domain

The cognitive domain seeks to make the thinking skills and intellectual processes required for the key stage 2 science sampling test explicit. Each question will be rated against the strands of the cognitive domain listed in tables 7 and 8 below.

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

The cognitive domain for science was based on Bloom's Cognitive Domain[2] . This provides a broad framework which the facts, processes and interpretations that need to be tested in science can be matched to. It allows broad testing of different aspects of children's scientific knowledge, including basic recall (eg of terminology); interpretation of data; and formulation of original responses to novel scenarios which draw on a broad scientific knowledge and skills base. Items have been grouped as follows:

- Items which assess factual knowledge (Knowledge and Comprehension)
- Items that require children to deduce an answer from presented data, or produce an answer based on use of scientific knowledge in a familiar or straightforward context (Application and Analysis)
- Items that require children to select which knowledge is appropriate, and possibly produce an answer that is original and tailored to a specific scenario (Synthesis and Evaluation)

The strands used to determine cognitive domain are:

- Complexity
- Response strategy

Items will be written for each strand of the programme of study as outlined in the content domain and rated according to the two strands above. As far as possible all strands of the programme of study will be assessed at each rating for both strands.

Descriptions of the two strands that will be used from 2016 are provided below. Tables for each dimension indicate how items will be rated, summarising the key points for each rating.

2 Bloom B.S., Engelhart M.D., Furst E.J., Hill W.H., and Krathwohl D.R. (1956) *'Taxonomy of educational objectives: the classification of educational goals; Handbook I: Cognitive Domain'* New York, Longmans, Green, 1956.

5.1 Complexity rating

The complexity scale described below is based on Bloom's cognitive domain. Knowledge is considered to produce the lowest level of cognitive demand, since it only involves children having to recall information that they already know. Evaluation has the highest demand, requiring children to demonstrate complex thought processes in order to put forward and argue their reasoning, supported by their knowledge and understanding. The cognitive levels have been grouped together in table 7 to give three ratings for complexity (from low to high).

Within the science sampling matrix as a whole, items will cover each rating for cognitive demand. However, some strands of the programme of study will lend themselves to being assessed at particular ratings for cognitive demand.

Table 7. Cognitive complexity rating scale

Item Dimension - Complexity	Knowledge and comprehension (low)	Application and analysis	Synthesis and evaluation (high)
The complexity of each component operation or idea and the links between them, based on Bloom's cognitive domain.	Remembers previously learned information and demonstrates an understanding of the facts.	Applies knowledge to actual situations, breaks down information into simpler parts and finds evidence to support generalisations.	Compiles component ideas into a new whole or proposes alternative solutions. Makes and defends judgements based on evidence.
	Recalls or describes simple factual information as required. Makes straightforward observations of features or objects. Reads or extracts information from simple data sources or text or diagrams.	Applies knowledge to given contexts. Gives simple explanations. Identifies patterns in data and makes comparisons. Makes predictions based on data given. Analyses data sources.	Makes links between different sources of evidence. Makes inferences and deductions from information given and own knowledge. Draws conclusions from evidence and relates to scientific understanding.

5.2 Response strategy rating

The response strategies are grouped into three ratings for science.

- Selected response items - these require the lowest level of cognitive demand as the way children must respond and the information they need in order to answer the question is given to them.
- Short constructed response items - these items require children to give some thought to the information required, but less about how they need to structure this information to give a response. Responses include plotting graphs and charts and drawing arrows on diagrams, where the instruction for what to do is provided but the child is required to think about how to display this information to demonstrate their scientific understanding.
- Extended constructed response items - these require the highest level of cognitive demand as children have to decide which information they will need to answer a question and how they are going to structure this information in their response.

Table 8. Response strategy rating scale

Response strategy	1. Selected response (low)	2. Short constructed response	3. Extended constructed response (high)
The extent to which a candidate is required to devise a strategy for tackling and answering an item.	The response is given in the question information and the strategy for responding is clearly provided.	The response is not given in the question information but the strategy for how the child is required to respond is clearly provided.	Children are required to construct their own extended response.
	Multiple choice, matching item, true / false style, circle, sequencing, ticking cells in a table, selecting correct data from a table, identifying from a key.	Short open response items, eg naming processes, parts, objects, variables. Describing, giving one difference, way or reason, completing bar or line graphs.	Longer open response items, that are likely to be two lines of writing, eg describing relationships or providing explanations.
		Drawing or completing diagrams by drawing arrows, circuit diagrams.	

6. Test specification

This section provides details of each test component and the sampling matrix design as a whole. Initially, the test will be administered on paper by external administrators and externally marked.

6.1 Summary of test and matrix design

Overall, a question pool totalling 330 marks will be selected. These questions will be divided into five papers for each content area (biology, chemistry and physics), giving a total of 15 test papers worth 22 marks each.

Each child will sit one paper from each content area, giving a total of three papers worth 66 marks.

Children will be given 25 minutes to complete each paper. The total testing time will be 75 minutes.

Table 9. Overall format of the item pool

Content area	Number of marks	Number of papers
Biology	110	5
Chemistry	110	5
Physics	110	5
Total	**330**	**15**

Table 10. Format of the test for each child

Component	Description	Number of papers	Number of marks	Approximate timing of component
Test paper b	Questions in a biology context	1(from 5 versions)	22	25 minutes
Test paper c	Questions in a chemistry context	1 (from 5 versions)	22	25 minutes
Test paper p	Questions in a physics context	1 (from 5 versions)	22	25 minutes
	Total	**3**	**66**	**75 minutes**

Each child takes three papers, one biology, one chemistry and one physics. Table 11 shows an example of how the different papers will be combined together within each of the 15 possible test combinations a child could be given: the numbers in the cells represent the paper number and the letter in brackets represents the component (biology, chemistry or physics).

Table 11: Matrix design

Combination	Order of administration		
	1st	2nd	3rd
1	1 (b)	6 (c)	11 (p)
2	2 (b)	7 (c)	12 (p)
3	3 (b)	8 (c)	13 (p)
4	4 (b)	9 (c)	14 (p)
5	5 (b)	10 (c)	15 (p)
6	6 (c)	13 (p)	2 (b)
7	7 (c)	14 (p)	3 (b)
8	8 (c)	15 (p)	4 (b)
9	9 (c)	11 (p)	5 (b)
10	10 (c)	12 (p)	1 (b)
11	11 (p)	3 (b)	10 (c)
12	12 (p)	4 (b)	6 (c)
13	13 (p)	5 (b)	7 (c)
14	14 (p)	1 (b)	8 (c)
15	15 (p)	2 (b)	9 (c)

This design may be adapted in each assessment year in order to include other papers of questions for specific purposes, such as TIMSS materials and trial material. These materials will form additional papers which will be added to the design.

6.2 Breadth and emphasis

The content and cognitive domains for the science tests are specified in sections 4 and 5. Although each element may not be included within each set of booklets administered to a child, the full range of content detailed in this document will be assessed over the full range of booklets. 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 in each test will be placed in approximate order of difficulty. Items within questions are not necessarily in order of difficulty as they need to fit within the context and show a follow through for a storyline (eg an investigation from planning to conclusion). Where possible, however, easier items are positioned at the start of a question to lead children in.

The following sections show the proportion of marks attributed to each of the areas of the content and cognitive domains, both in the item pool as a whole and in the test administered to each child.

6.2.1 Profile of content domain

Questions are designed to be related to one of the three content areas of biology, chemistry and physics. Table 12 shows the proportion of marks in the item pool expected to be related to each of the three core content areas. Each content area will contribute one third of the marks in the item pool.

Table 12: Profile of marks by content area

Content area	Number of marks in item pool	Number of papers in item pool	Number of marks per paper
Biology	110 (33%)	5	22
Chemistry	110 (33%)	5	22
Physics	110 (33%)	5	22
Total in test	330	15	66

In 2016, only elements that are common to the old and new curriculum documents will be assessed from the year 3-4 content areas.

Within each content area, some items only assess the related strand (in the national curriculum), while other items assess the related strand plus 'Working scientifically'. 'Working scientifically' will contribute between 25 and 35 per cent of the item pool. In order to ensure this is balanced across the matrix design, each paper will be constructed to contain a minimum of five marks assessing 'Working scientifically', meaning that each child's test will contain a minimum of 15 marks addressing this area of the content domain.

Table 13: Profile of marks assessing 'Working scientifically'

Number of marks in item pool	Minimum number of marks per paper
83–115 (25%–35%)	5

6.2.2 Profile of cognitive complexity

Items will be classified according to the cognitive complexity strand of the cognitive domain (see section 5). Table 14 shows the expected proportion of marks classified at each level of cognitive complexity within the item pool as a whole. No specific restrictions are placed on the distribution of marks within each individual paper but it is intended that each paper covers a range of items at different levels of complexity.

Table 14: Profile of marks by cognitive complexity

Cognitive complexity	Number of marks in item pool	Percentage of marks in item pool
Knowledge and comprehension	99-165	30-50%
Application and analysis	99-165	30-50%
Synthesis and evaluation	33-99	10-30%

6.2.3 Profile of response strategy

Items will also be classified according to response strategy (see section 5). Table 15 shows the expected proportion of marks classified at each level of response strategy within the item pool as a whole. No specific restrictions are placed on the distribution of marks within each individual paper but it is intended that each paper covers a range of different types of items.

Table 15: Profile of marks by response strategy

Response strategy	Number of marks in item pool	Percentage of marks in item pool
Selected response	132-165	40-50%
Short answer	83-115	25-35%
Extended response	66-99	20-30%

6.3 Format of items and responses

All three types of paper (biology, chemistry, physics) contain the same range of item types.

Selected response items, where children are required to select a response from a limited range of options, including:

- multiple choice items
- matching items
- true / false items.

Short answer items, where children are required to produce a short response and no options have been given. Examples of short answer items include:

- naming an object or process
- completing a sentence with a missing word
- completing a table or chart
- labelling a diagram or key.

Extended response items, where children are required to construct their own longer response. Examples of extended response items include:

- explanations
- justifications for previous answers
- various elements of planning an investigation
- graphing or describing a data set.

6.4 Marking and mark schemes

The test will be externally marked on-screen by trained markers.

The tests will be administered securely and kept confidential. Mark schemes will be developed alongside the test materials so that the tests can be marked reliably, however they will not be published.

In each year the science sampling tests are administered in schools, a sample of questions will be published with an accompanying mark scheme.

The mark scheme 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. Additional guidance will indicate minimally acceptable responses and unacceptable responses, where applicable.

For all questions, the mark schemes will be developed during the test development process and will combine the expectations of experts with examples of children's responses that have been 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:

- A content domain reference
- Marks – the maximum number of marks for each item in a question.
- Requirements – for open response items this will include a generic statement describing the type or types of responses that will enable a child to gain credit, followed by a number of exemplar responses to indicate the types of answer a child may write. For selected response items this will indicate the response required for the child to gain credit.

- Allowable responses – this section will provide exemplars of allowable responses. Allowable responses are minimally acceptable and are usually not as good from a scientific perspective.
- Additional guidance – this section will give information on response types that are not creditworthy because they are either insufficient or incorrect.

6.5 Reporting

Analysis will be undertaken using the combined data from all children in the sample in order to produce a scaled score representing the performance of the sample as a whole. The purposes of translating raw scores onto scaled scores for science sampling are twofold:

- Performance can be reported as a whole, having taken into account the varying difficulties of the specific test combinations taken by different children.
- Scaled scores retain the same meaning from one year to the next so a particular scaled score reflects the same level of attainment in one year of administration as in the previous, allowing national performance to be compared across years.

Additionally, an estimate of the proportion of children who have achieved the required standard on the test will be produced. A standard-setting exercise will be conducted on the first live test in 2016 in order to determine the scaled score needed for a child to be considered to have met the standard. This process will be facilitated by the performance descriptor in section 6.6 which defines the performance level required to meet the standard. In subsequent years, the standard will be maintained using appropriate statistical methods to translate performance across a new item pool into scaled scores. The scaled score required to achieve the expected standard will always remain the same.

The exact scale for the scaled scores will be determined following further analysis of trialling data. This will include a full review of the reporting of confidence intervals for scaled scores, as well as other scale properties.

6.6 Desired psychometric properties

The focus of the outcome of the science sampling model is to provide an estimate of the proportion of children nationally who would have achieved the required standards. While it is important to minimise the standard errors of measurement throughout the scale, it is particularly important to do so around the threshold of the expected standard.

6.7 Performance descriptor

This performance descriptor describes the typical characteristics of children whose performance is at the threshold for the expected standard at the end of key stage 2. Children 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 science. This performance descriptor will be used by teachers to set the standards on the new tests following their first administration in 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 pen-and-paper test (see the content domain in section 4).

The complete suite of key stage 2 science sampling tests taken in each live sample year include questions assessing the entire key stage 2 programme of study in the national curriculum including the 'working scientifically' strand. Each child will only take three tests. It is therefore the aggregated results from the complete test suite that will provide an indication of how children are performing in science across England. It is assumed that children working at the threshold will be generally secure in the scientific knowledge and skills they acquired at key stage 1.

6.7.1 Overview

Children 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 either working at the threshold of the expected standard or if the question is assessing a topic area they have not revisited in primary school for several years.

Questions will range from those requiring recall of facts to those requiring synthesis and evaluation. There will be a variety of question formats including selected response, short answer and constructed response where no strategy is provided within the question.

Question difficulty will be affected by the scientific content (including the 'working scientifically' strand) and by the different strands of the cognitive domain, such as question complexity and the extent to which support is given in the question to help children identify or construct their response. This should be borne in mind when considering the remainder of this performance descriptor. Children working at the threshold of the expected standard may not give correct responses to all questions, especially if there is a particularly high cognitive demand. This will be true even when the performance descriptor determines that a skill or knowledge should be within the child's capacity if working at the expected standard.

The following sections describe examples of the threshold standard, some of which a typical child in Year 6 should be able to demonstrate. It is recognised that different children will exhibit different strengths, so this is intended as a general guide rather than a prescriptive list. It is also recognised that the topics in the science national curriculum are described by year group, therefore a topic covered in a discrete year may not be covered again at key stage 2.

The content in the national curriculum has been divided up into 'Working Scientifically' and the three scientific disciplines. The performance descriptors below reflect this.

6.7.2 Working scientifically

Children should be able to recall, use and apply their knowledge, understanding and skills in 'working scientifically' to familiar and unfamiliar contexts. Children should be able to (with no order of priority or hierarchy intended):

- recall and use appropriately terminology such as accurate, conclusion, evidence, fair test, method, observe, pattern, prediction, reliable, results, supports (evidence) and variable
- for a given task they can identify the most appropriate approach for answering scientific questions and select the most appropriate equipment and sources of evidence needed for a task

- plan different types of scientific enquiry, make careful observations, take accurate measurements or readings using the appropriate units as required and identify when to repeat measurements, if necessary, to ensure given results are reliable
- record, present and interpret data from different sources, using a range of methods, including tables, graphs (bar charts and line graphs), diagrams and keys
- apply their understanding of scientific concepts to draw valid conclusions from data
- use data to make predictions for missing values
- identify or use evidence to support or refute ideas or arguments
- recognise the validity and reliability of evidence and the difference between fact and opinion.

6.7.3 Biology

Children should be able to recall, use and apply their knowledge and understanding to familiar and unfamiliar contexts. They should draw conclusions, make simple evaluations and synthesise information.

Children should be able to (with no order of priority or hierarchy intended) respond to the majority of statements below, but may be less secure in areas that were covered in Years 3 or 4:

- recall and use appropriately terminology such as adaptation, circulatory system, classification, consumer, evolution, function, germination, invertebrates, nutrients, pollination, predator, prey, producer, reproduction, seed dispersal and vertebrates
- describe the processes involved in different stages of the flowering plant's life cycle and the function of different parts of flowering plants
- describe how water and nutrients are transported in plants
- compare the requirements of plants and animals to live and grow well
- compare the similarities and differences between the life cycles of different animals (including humans and other mammals, birds, amphibians, and insects)
- describe the functions of parts of the digestive system in animals
- describe the functions of the main parts of the circulatory system (including the transport of nutrients and water) in animals
- describe the functions of the skeleton and muscles in animals
- describe the effects of diet, exercise, drugs and lifestyle on how our bodies function in the long and short term
- construct and interpret food chains
- use keys to group, classify or identify living things, and construct simple dichotomous keys
- describe the main characteristics used to group plants, animals and micro-organisms according to the main groups (vertebrates, invertebrates, birds, mammals, reptiles, fish and amphibians) in the classification system
- explain how a change in an environment may have an impact on living things

- identify that there is variation between offspring and between offspring and their parents because of differences in inherited characteristics
- describe how plants and animals have adapted to their environment and how this may have led to their evolution
- describe how living things have changed over time and that fossils provide information about living things in the past.

6.7.4 Chemistry

Children should be able to recall, use and apply their knowledge and understanding to familiar and unfamiliar contexts. They should draw conclusions, make simple evaluations and synthesise information.

Children should be able to (with no order of priority or hierarchy intended) respond to the majority of statements below, but may be less secure in areas that were covered in Years 3 or 4:

- recall and use appropriately terminology such as condensation, °C (degrees Celsius), evaporation, filtering, freezing, insoluble, melting, mixture, non-reversible, properties, reversible, solidifying, soluble and solution
- compare the characteristics of different states of matter (solids, liquids and gases)
- describe how materials can change state with reference to temperature, and explain everyday phenomena (including the water cycle) where changes of state occur
- classify and group materials according to properties such as appearance (for rocks), hardness, solubility, transparency, conductivity and magnetism
- describe the advantages and disadvantages for the uses of everyday materials based on an understanding of their properties using appropriate terminology
- identify and recognise everyday phenomena where dissolving occurs
- describe how to appropriately separate different mixtures of materials, including solutions
- identify and compare reversible and non-reversible changes
- describe in simple terms how fossils are formed
- describe the composition of soil.

6.7.5 Physics

Children should be able to recall, use and apply their knowledge and understanding to familiar and unfamiliar contexts. They should draw conclusions, make simple evaluations and synthesise information.

Children should be able to (with no order of priority or hierarchy intended) respond to the majority of statements below, but may be less secure in areas that were covered in Years 3 or 4:

- recall and use appropriately terminology such as air resistance, attraction, conductor, friction, gravity, insulator, newtons (N), opaque, orbit, pitch, repulsion, sphere, translucent, transparent, vibration, voltage, volume and water resistance

- explain how we see other objects (from a single reflection) and represent this in simple diagrammatic form
- explain shadow formation and how the size of shadows may change
- explain how sounds are made and describe how they require a medium to travel through from the source to the ear
- describe how volume can be changed with reference to vibration
- describe how the features of an object determine the pitch of a sound
- describe the shape of bodies (spheres) in the solar system and the movement of bodies in the solar system relative to each other
- explain how day and night, including the apparent movement of the sun across the sky, are related to the Earth's rotation
- draw or complete a simple series circuit diagram using recognised symbols including straight lines for wires
- explain how changes made to a circuit can affect how it works
- identify and describe the effects of contact and non-contact forces on moving and stationary objects
- describe the effects of magnets on magnets and other materials
- describe how simple pulleys, levers, springs and gears increase the effects of a force.

7. Diversity and inclusion

The Equality Act 2010 sets out the principles by which the national curriculum assessment and its 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' (Ofqual, http://www.ofqual.gov.uk/files/2011-regulatory-framework-for-national-assessments.pdf).

The end of key stage 2 science sampling test should:

- use appropriate means to allow all children to demonstrate their knowledge and skills in biology, chemistry and physics
- provide a suitable challenge for all children and give every child the opportunity to achieve as high a standard in science as possible
- provide opportunities for all children to achieve, irrespective of gender, including children with special educational needs, children with disabilities, children from all social and cultural backgrounds and those from diverse linguistic backgrounds
- use materials that are familiar to children and for which they are adequately prepared
- not be detrimental to children's 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' (Ofqual, 2012; www.ofqual.gov.uk/news/new-language-accessibility-guidance-published).

In order to improve general accessibility for all children, where possible, questions will be placed in order of difficulty. Accordingly, to be consistent with all national curriculum tests, attempts have been made to make the question rubric as accessible as possible for all children, 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 children, and taking into account feedback from expert reviewers. Within questions, items are not necessarily in order of difficulty because they need to fit within the context and show a follow through for a story line (eg an investigation from planning to conclusion). Where possible, however, easier items are positioned at the start of a question to lead children in.

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 children 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 relevant ARA will be available to eligible children as required.

Appendix: Glossary of terminology used in the test framework

cognitive domain	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. The cognitive domain, as shown in this framework, also identifies other factors that may influence the difficulty of the questions.
component	A section of a test, presented to children as a test paper or test booklet. Some tests may have two or more components which each child needs to sit in order to complete the test. The key stage 2 science sampling test comprises three components.
content domain	The body of subject knowledge to be assessed by the test.
construct irrelevant variance	Construct irrelevant variance is the variation in children's test scores that does not come from their knowledge of the ideas being tested. It can result in children gaining fewer marks than their knowledge would suggest or lead to the award of more marks then their knowledge alone would deserve. 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 children to answer correctly without having the required subject knowledge.
distribution	The range of possible scaled scores.
domain	The codified definition of a body of skills and knowledge.
item	The smallest unit that can be awarded a mark within a science question.
mark scheme	The document explaining the creditworthy responses or the criteria that must be applied to award the mark for a question in the test.
matrix sampling	A sampling assessment whereby different groups of children take different test instruments, which are all linked together in a matrix.
national curriculum	For each subject and key stage, the national curriculum outlines the content and skills that should be taught in schools.
performance descriptor	Description of the typical characteristics of children 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.

programme of study	The statutory curriculum of subject knowledge, skills and understanding for a key stage. The key stage 1 and 2 programmes of study are published online at www.education.gov.uk/schools/teachingandlearning/curriculum.
question	A group of related items assessing a common context.
raw score	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 child 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.
sample	The group of children selected to take the test in a given year.
scaled score	A score which has been translated from a raw score onto a score on a fixed, defined scale. This allows performance to be reported on a consistent scale for all children, 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 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	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 child had that child taken the test repeatedly under the same conditions.
standard setting	The process of applying the standard onto a particular test in order to determine the score required for a child 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 contains a test specification.
test specification	A detailed specification of what is to be included in a test in any single cycle of development.
truncate	To shorten by removing ends.

Standards & Testing Agency

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 science sampling. It may also be of interest to schools with children 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

Related information

Visit the Department for Education's website at www.education.gov.uk/ks2 and www.gov.uk/government/collections/national-curriculum for all related information.

For more copies

Printed copies of this document are not available. It can be downloaded from the Department for Education's website at www.education.gov.uk/ks2.